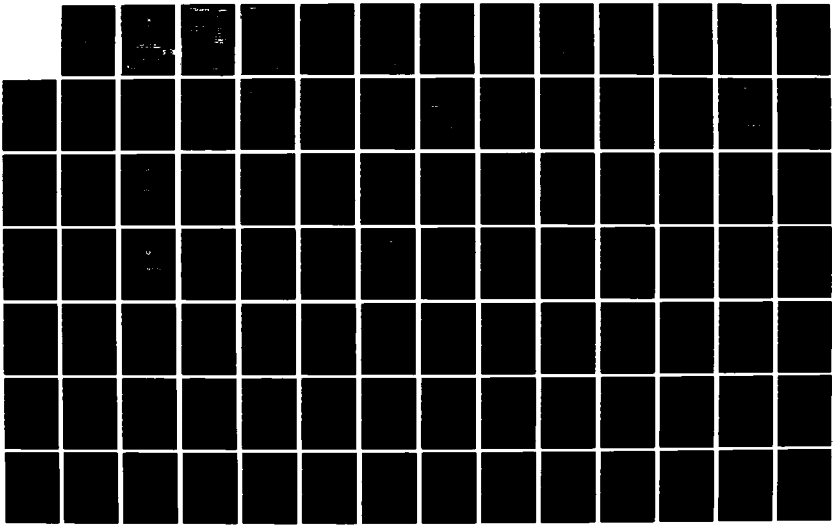
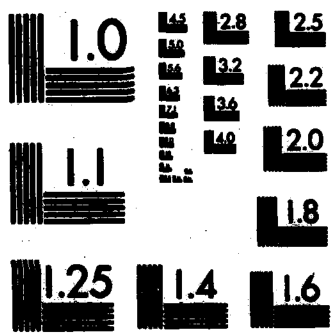
AD-A121 990 QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS J E ENGEL APR 80 AFIT-LS-32 1/7

UNCLASSIFIED F/G 15/5 NL





(2)

QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY (ATC)
AIR FORCE INSTITUTE OF TECHNOLOGY
SCHOOL OF SYSTEMS & LOGISTICS
Wright-Patterson Air Force Base, Ohio

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER LS-32	2. GOVT ACCESSION NO. AD-A121990	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER		5. TYPE OF REPORT & PERIOD COVERED LS-32 Textbook
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Edited by John E. Engel, Lieutenant Colonel, USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Systems and Logistics Air Force Institute of Technology, WPAFB OH		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Department of Communication and Humanities AFIT/LSH, WPAFB OH 45433		12. REPORT DATE April 1980
		13. NUMBER OF PAGES 644
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE 20 OCT 1982

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE AFR 190-1

Approved for public release; distribution unlimited

Director of Public Affairs

Air Force Institute of Technology (ATC)
Wright-Patterson AFB, OH 45433

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Problem-solving techniques
Quantitative Tools
Logistics
Decision Making
Statistics

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The increasing cost and complexity of our defense force combined with frequent reductions in resources requires each logistics manager to become knowledgeable of the quantitative problem-solving techniques available. This text is designed to serve as a ready reference of those quantitative tools most frequently used.

LS-32

School of Systems and Logistics
Air Force Institute of Technology
Wright-Patterson AFB, Ohio 45433

QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER

Edited By

John E. Engel
Lieutenant Colonel, USAF

April 1980

PREFACE

This handbook is the first edition of a supplemental text to support the curriculum presented in the resident graduate management programs of the Air Force Institute of Technology. It is also intended to be used by the practicing logistics manager as a ready reference following a formal academic program.

The increasing cost and complexity of our defense force combined with frequent reductions in resources requires each logistics manager to become knowledgeable of the quantitative problem-solving techniques available. This text is designed to serve as a ready reference of those quantitative tools most frequently used.

Chapters 1 through 5 discuss concepts and techniques used in specific functional areas of logistics management. Chapters 6, 7, and 8 provide a brief review of basic statistics. The remaining 14 chapters cover specific quantitative tools used to aid the logistics manager in the decision-making process. No classified or copyrighted material is contained in this handbook.

The handbook is a compilation of the work of several authors. Each is identified in the Table of Contents. My sincere thanks to them for their effort. I also wish to thank and acknowledge Kathy Taylor, Shirley Sawyer, and Peggy Limbach for their handling of all the typing and numerous other details in a prompt and accurate manner.

April 1980

JOHN E. ENGEL, Lt Col, USAF
EDITOR

TABLE OF CONTENTS

CHAPTER

CHAPTER 1

SUPPLY

Over the years our Air Force Major Commands and base supply units have developed numerous management indicators to assist in guiding their efforts toward improved mission support. Many of these indicators have been modified due to changes in computer programs and management needs. Frequently, officers new to the Air Force and officers from the rated career field are assigned to base level supply management positions. They may find they have a lack of understanding of the supply "business" and need information on which to base management decisions. This chapter was developed to assist the officers new to the supply field to understand some selected supply management indicators. With this understanding, the officers should be capable of making rational decisions earlier in their supply assignment and become a more effective contributor to the supply support mission.

Our inventory management actions relate to managing the trade-offs in costs for holding inventory, costs of ordering inventory and out of stock costs. We must effectively issue items requested from stock and if no stock exists, we must have a timely ordering system to respond to the customer needs. For the most part, managers at all levels no longer expect the supply system to meet every need with a 100% issue rate from stock. These managers realize the cost of holding stock is of such magnitude that item stockage occurs only when justified by past demands or when justified by special projects, programs or mission changes.

Base level supply is essentially the retail level in our distribution system. It is a node in a larger system with supply sources of DLA, Army depots, Navy depots, GSA, and Air Force depots being the wholesale level. The objectives of these different levels of supply are essentially the same, maximum customer support while considering the cost tradeoffs of inventory management and minimizing the total backorder days when obtaining items not in stock. Backorders will occur because it is not physically and economically possible to hold in warehouse inventory all possible customer requirements.

Managing the supply system requires management attention in two areas. Control, issue and replenishment of warehouse inventory and requisition control and issue of customer needs for items not in stock. The use of management indicators in these two areas is extremely helpful in determining the degree of success or failure of the distribution system in response to customer demands. The management indicators should be derived from data sources that are readily available, present valid data, be limited to those indicators that are necessary, and,,of course, be understood and easily communicated. Ten indicators frequently used by Major Commands and bases were selected for analysis. The indicators are as follows:

1. Customer Support Effectiveness

2. Stockage Effectiveness

3. Not Mission Capable Supply

4. Priority Requisition Rate

5. Bench Stock Support

6. Overall Inventory Accuracy

7. Delinquent Document Rate

8. Unauthorized Equipment

9. Serviceable Balance - No Warehouse Location

10. Repair Cycle Support (Delayed Repairs)

Each indicator will be examined in the following pages and specific questions will be answered. Why is the indicator important? How is the indicator determined? What management actions should be taken to improve an unfavorable indicator trend?

1. CUSTOMER SUPPORT EFFECTIVENESS

IMPORTANCE OF INDICATOR: Customer support is the sole reason for the existance of a Base Supply. The support, however, must be provided within the tradeoffs previously mentioned and the numerous constraints imposed by higher level supply directives.

This indicator shows a percentage of issues made in relation to total requests for items that have been assigned stock levels. It does not refer to data analysis of requisitions for items that did not have a stock level. Past demands for an item have justified the need for a stock level or a special level may have been justified and established. In either case, a stock level existed and a balance was not available for issue, thus making a back-order necessary. This situation causes two requisitions for the same line item. One requisition for the customer and the requisition that already existed for stock replenishment. It is also possible that the requisition for the customer could be in a requisition Priority Group 1 or 2 and therefore add to the numerous priority requisitions within our Uniform Material Movement and Issue Priority System. Customer support effectiveness directly reflects the ability of supply managers to manage warehouse inventories. It is obvious that low customer support effectiveness may degrade an organizations capability to support its mission.

DATA SOURCE: Customer support effectiveness data is obtained from the UNIVAC 1050-II M-32 (Monthly Base Supply Management Report). This report breaks out each type organization (AFM 66-1, Civil Engineers, Vehicle Maintenance, and others) separately as well as provides an overall percentage of customer support effectiveness. The M-24 (Organization Effectiveness Report) breaks out each organization code separately and allows the review of separate shop or branch codes within an organization.

FORMULA:

$$\text{Customer Support Effectiveness} = \frac{\text{Total issues from authorized stock}}{\text{Total issues from items authorized stock + due outs with due ins of items authorized stock}}$$

MANAGEMENT ACTIONS THAT CAN BE TAKEN TO REVERSE AN UNFAVORABLE TREND:
It is no small undertaking to improve an unfavorable trend of this indicator. The trend reflects a summary of thousands of variables within the supply system. It also reflects the support given to all organizations on a base, perhaps hundreds.

A good way to start in analyzing this indicator is to select an organization that is making a significant negative contribution and then review those items with the lowest issue rate. The M-24 report can be used to identify the organization. Identification of individual stock numbered items can be accomplished by asking the organization for a list of all System Requirement Designators (SRD's) they use when calling in requests to Demand Processing. Six to ten SRD's per organization is normally the maximum an organization will have. Review of the SRD's with the Base Supply File Maintenance Section is necessary to insure that the applicable SRD's are loaded into the computer data bank. After verification, request the Automatic Data Processing Equipment (ADPE) Unit to process the SRD's through the R-37 report (SRD Demand Data Analysis) using the current D-13 report tape (Daily SRD Update). The R-37 report will output cards for every stock number requested using the SRD's. The R-37 report cards can be input to the R-29 report (Problem Item List) to produce a computer printout for manual review of item history. A thorough examination of the data history (demands, levels, order and ship time) on each problem stock numbered item needs to be accomplished. As problems are found, positive long range solutions need to be implemented.

As you can see, improving customer support for stocked items is a time consuming task. The analysis of the problem may require numerous hours of review for each organization. Even though locating the problems is difficult, developing long range solutions and applying quality control management to the solutions will be even more time consuming and difficult. The procedures described are extremely effective. At one Air Force Base a stockage effectiveness rate of 65% was improved to 87.79% in only seven months by use of these procedures.

The base supply manning is not sufficiently large to implement the above suggestions for managing all items that degrade customer support. An over aggressive effort to improve this indicator will divert subordinates attention from other duties and negatively affect the supply account in other areas.

2. STOCKAGE EFFECTIVENESS BY WEAPON SYSTEM

IMPORTANCE OF INDICATOR: Most of our bases have a flying mission and communications equipment that require supply support. These weapon systems are not only high cost but are also critical to the defense of the United States. Monitoring an indicator that follows weapon system support is helpful in knowing the success of supply to support these essential requirements. In weapon systems support, the number of supply sources (depots) is limited. This permits our supply personnel to concentrate management efforts on a few depots and obtain extremely helpful computer reports advising of depot support.

DATA SOURCE: The M-11 (Weapon System Support Effectiveness Report) provides the data for this indicator.

FORMULA: Percentages are computed by dividing line items issued plus line items backordered minus line items B/O not authorized stock into line items issued.

MANAGEMENT ACTIONS THAT CAN BE TAKEN TO REVERSE AN UNFAVORABLE TREND: The stockage effectiveness indicator follows the percentage of total requests for items that were filled from base stocks. The indicator reflects overall weapon system support and permits management to isolate sources of supply that are contributing to good as well as poor support. Individual stock numbered items causing low effectiveness can be isolated and thoroughly examined. Examination would consist of analysis of order and ship time, followups, reject rate, use of part number loaded in the computer, receipt analysis, demand levels, and perhaps analysis of some mission change that might have occurred causing a variance in demand patterns. In analyzing order and ship time it is important to know how often the Automated Data Processing Equipment (ADPE) Unit is releveling. The M-32 report may show 30 to 50 times a month but this could be inflated by the ADPE Unit releveling 5 to 10 times on slow days. The ADPE Unit has a requirement to complete releveling each computer date. Daily accomplishment of releveling is very important since the urgency of need "B" and "C" requisitions (includes stock replenishment) account for as much as 90% of all requisitions placed. Proper releveling actions lead to building stocks and therefore permits a higher percentage of issues. When issues are made, customers will have fewer cancellations. Cancellations are another area rated in supply management.

Requisition follow-up is another important factor in stock management. The ADPE Unit is required to complete requisition follow-up at least twice a month. If both follow-ups are completed on the same day and early in the month, it is obvious that valuable valid status will not be available to supply managers. When using this method, many requests will not meet the time frame for follow-up and consequently they will be overlooked for as long as 30 days. The Stock Control Unit will be forced to use manually prepared follow-up documents which require numerous manhours. With proper management, the computer could have accomplished the follow-ups with no

manhour expenditure. Follow-ups should be accomplished the first and third week of each month. If time is available, follow-ups could be completed daily just after releveling is complete. This action will provide the Stock Control Unit with the most current status each day. The computer program will not allow follow-up on the same item every day. Follow-up will be accomplished only on those items that are past due the follow-up time frame.

Computer inputs that add to the "Reject Rate" may also negatively impact on Stockage Effectiveness. Rejects that are not promptly corrected and re-input cause delays in the order and ship time. Therefore, rejects must be cleared in a timely manner to avoid degrading stockage effectiveness.

A significant improvement in use of the UNIVAC 1050-II capabilities has been the addition of item part numbers to item records. Nearly every National Stock Number has a part number and loading the part number in the computer greatly reduces requisition processing time. With the part number loaded, maintenance personnel can call in their request by part number. The computer converts the part number to a stock number and causes an issue or backorder. The key to this process is to correctly load the part number, and Technical Order (T.O.) figure and index. After the loading of an items data is complete, manual research for requisitions will no longer be required. All bases may not be taking full advantage of the computer's capability. An aggressive program to load part number data will yield immediate and long-term customer support benefits. Delays in requisition research and processing will be greatly reduced. An effective way to build a computer data base is to ask maintenance personnel to provide part number and stock number data at the time of request call-in. Initial building of the data base in this manner is slow but very rewarding in the long run.

Delays in receiving items for stock will, of course, also affect "Stockage Effectiveness". The M-32 report, Source of Supply Summary, Receipts Summary, will identify problem sources of supply. This report, along with the Q05 Routing Identifier Listing, will further identify supply sources that need to be reviewed. The R29, Problem Item List, can then be used to selectively obtain item record data from the computer. The M-32 and Q05 reports are not good tools to work with if the ADPE Unit is not processing status daily or the Receiving Section is behind in processing receipts.

In reviewing demand levels you may find maximum levels established by a depot Inventory Manager. This occurs because the quantity of a line item is critically short in the supply system. Because they are short, bases cannot be permitted to hold the high stock levels that demands may justify.

Aggressive management of this indicator will prove helpful in maintaining or achieving a high level of support. The importance of Weapon System stockage effectiveness cannot be overstated. Constant review, follow-up, and coordination of requirements are essential to maintain effective mission support.

3. NOT MISSION CAPABLE SUPPLY (NMCS)

IMPORTANCE OF INDICATORS: It is very important to know the down time of the major item in a weapon system due to the inability of the supply system to respond. When items are not immediately available for issue from warehouse stocks one of several actions would take place:

a. Review R26 Due In From Maintenance (DIFM) Listing and D19 Awaiting Parts (AWP) Validation Listing for status of item in DIFM.

b. Withdraw the item from the War Reserve Spares Kit if available.

c. Maintenance may cannibalize the required item from another major item.

d. Requisition the required item from the appropriate depot.

e. Locate the required item at another base and have the item shipped. This can be accomplished after receipt of delayed status on the requisition. Lateral support action causes double shipping costs, since in most cases the base that shipped the item must now also requisition their shortage.

One of these actions would eventually resolve the item shortage.

Another important factor relating to NMCS is the affect the priority requisitions have on the Uniform Materiel Movement and Issue Priority System (UMMIPS). If it becomes necessary to acquire the item from lateral support or a depot, a requisition in Priority Group 1 will be created, this high priority enjoys "immediate" communication and air shipment of the item. As the distribution system receives a higher percentage of these priority 1 requests, the system becomes more costly to operate and subject to request processing delays at all priority levels. NMCS rates are affected by new demands (items not authorized stock), the age of the weapons system, depot support and base supply management of items authorized stock.

DATA SOURCE: Data for this indicator is obtained from listings maintained by the Materiel Control Unit of the maintenance activity and NMCS data identified in D23 Mission Capable (MICAP) Report.

FORMULA: $$\text{NMCS Rate} = \frac{\text{Number of NMCS Hours}}{\text{Number of Available Weapon System Hours}}$$

The key factor is "NMCS hours" rather than "NMCS occurrences". Once the initial NMCS request is unsatisfied, the rate is dependent upon the ability of the distribution system to react to the NMCS requisition.

MANAGEMENT ACTIONS THAT CAN BE TAKEN TO REVERSE AN UNFAVORABLE TREND: A NMCS rate will exist because it is not economical to have stock levels for every item that will require repair or replacement. However, improving an unfavorable trend will require agressive continuous work. In determining the reason for out of stock situations, there is frequently a requirement for extensive review of supply data and constant follow-up by supply personnel. Each requisition for an item to resolve a NMCS requires constant monitoring. In addition to monitoring existing NMCS conditions, supply personnel must determine if the NMCS rate is so high that a general review of weapon system stock levels, inventory procedures, order and ship time or other supply management areas require analysis to identify shortcomings in the system.

Often times problems can be found in computing the NMCS rate. Therefore, coordination with maintenance may be necessary to be assured that the "Number of Available Weapon System Hours" are correct and that NMCS start and stop times are being logged in correctly. Maintenance can also be of considerable assistance in reducing the NMCS rate by performing cannibalization action when appropriate.

4. PRIORITY REQUISITION RATE

IMPORTANCE OF INDICATOR: As you would expect, priority requisitions within our logistics distribution system receive more management attention than routine requisitions. This is because of the procedures established by the Uniform Materiel Movement and Issue Priority System (UMMIPS). UMMIPS establishes criteria whereby requisitions can be categorized and processed according to a priority that is relative to the requesting activities' military mission importance.

DOD governs the requisition, issue, and movement of all materiel managed by its separate service components, and through agreement, those materiel items furnished to the components by General Service Administration (GSA). DOD has also directed that each component maintain a priority system as a management tool for the proper allocation of materiel resources within criteria that has been uniformly established.

DOD has prescribed five categorizations and the criteria for assigning each. A categorization is called a Force/Activity Designator (FAD) and identifies all U. S. and certain foreign assistance pact countries as to their respective places in the spectrum of military essentiality.

While the FAD is very important in the establishment of priorities it is but one of two factors; the other being the "urgency of need" designator (UND). The UND is a means by which a requesting organization can convey the degree of urgency of its materiel needs. When the requesting organization's FAD is combined with the UND it assigns a materiel requirement a "Priority Designator" (PD).

Priority Designators are a series of numeric codes from 01 to 15 that are used to categorize the relative importance and urgency of requests. When the requests are categorized and refined through a base supply requisitioning program, requisitions are prepared to the higher level supply source. It should be noted that the MILSTRIP/UMMIPS criteria apply only to those requisitions that leave the base to the Inventory Manager (IM) System Manager (SM), DLA, GSA or another military department.

The requisitions are segregated into three priority groupings to facilitate handling and processing. Each group has established guidelines and time limits for every step in the requisition processing. Priority groupings govern the type of communications media and priority used for requisition, status information, follow-ups and replies; the speed at which the supply source processes the requisitions; the materiel selection and issue criteria; and the selection of the mode of transportation used to ship the materiel to the requesting agency.

Selection of the proper UND is the responsibility of the Commander of the requesting activity. By conscientiously weighing UND alternatives, individuals can insure that demands which have the most serious impact on mission capability will receive priority processing and handling through the entire requisition cycle. The UMMIPS is designed for selective use of priorities based on predetermined factors. Automatic and uncontrolled assignment of priorities is an abuse of the system, and as such cannot be condoned.

The wrong UND, once assigned by the user, becomes the prime contributor to the abuse of the system. Whenever the user, who could very likely be a mechanic on the line, determines how detrimental the lack of an item is to the operational support of a system, a sequence of significant events is started.

After approval to requisition an item and the priority has been certified, the requisition leaves the base by high-priority communication; the depot processes the request, takes the needed item from stock and packs the order for shipment; and the transportation section ships the materiel by the most expeditious method consistent with the priority group. Each of these actions is extremely costly, but considered worth the expense to get the needed item to the user in minimum time. If the priority is legitimate, the expense can be justified; if not, the assignment of the incorrect priority results in senseless expenditures, even if the mistake is made only once.

DATA SOURCE: The Priority Requisition Rate Data is identified in the M-32 Report for the month being reviewed. The Q05 (Routine Identifier Listing) adds each month together for a total FY 1 Oct - 30 Sep review.

FORMULA: Overall total Priority Group 1 and 2 requisitions and special requisitions divided by overall total Priority Group 1, 2, and 3 requisitions and special requisitions.

MANAGEMENT ACTIONS THAT CAN BE TAKEN TO REVERSE AN UNFAVORABLE TREND: All organization and unit commanders have an inherent responsibility to exercise the degree of supply management, surveillance, and discipline to enforce the concepts and procedures of the UMMIPS system. Without a high level of command interest, the logistics system cannot satisfactorily respond to the unit's legitimate materiel needs. Some of the steps the commanders can take to establish proper supply management procedures and to enforce the discipline are to:

(1) efficiently plan, forecast, and submit requests for the materiel needed with as much lead time as possible. This is basic to allowing the supply system to plan for support instead of reacting to an emergency.

(2) establish a system whereby each UND "A" request is properly reviewed and that the required delivery dates (RDD) and the quantity of materiel required reflect realistic and valid factors; and request only the quantity of materiel needed.

Although the responsibility of enforcing supply discipline is that of the unit commander, in the final analysis it is the Chief of Supply who must see the commander's directions are being followed.

One way the Chief of Supply can help to achieve correct adherence to UMMIPS procedures is by informing each of his customers of the restrictions and operating procedures that control the degree of supply support that can be given and the ways in which the customer can assist to improve the system's reaction. For example, once the customer understands the prevailing order and shipping time the more likely he is to submit requests on time. Further, the customer should be advised to avoid, or at least minimize, excessively high priority requisitioning. The highest degree of cooperative materiel planning and forecasting by the Chief of Supply and the customer is basic to maintaining a well balanced stock position. In the requests the customer submits to supply, realism must be reflected not only in the quantities but also in the urgency of need. The capability of the logistics system to react satisfactorily to all valid mission requirements depends largely on the validity of the priorities assigned by the customer. When the majority of the requisitions within the distribution system are priority, then essentially all requisitions become routine.

5. BENCH STOCK DUE OUTS

IMPORTANCE OF INDICATOR: The importance of the bench stock system is often not understood by supply people. A maintenance weapon system bench stock is the "life blood" of the base level repair cycle. Base maintenance facilities are the best source of supply for major end items of a weapon system. You can expect that a good base maintenance operation will repair 70% to 90% of all repair cycle items and return them serviceable to base supply warehouse stocks. Base supply has a critical input to this repair cycle. They must make available to maintenance those small, usually low cost, high consumption items that are necessary to make end item repairs. When stocks are not available, the repair of the major end items will be

delayed. The item may be sent to the depot for repair causing added cost and the flow of an additional item in the distribution system. It is also obvious that an efficient repair cycle at base level will reduce the NMCS rate because more serviceable assets will be available for immediate issue to customers.

DATA SOURCE: The "Bench Stock Summary" page of the M-32 provides this data.

FORMULA: Bench Stock Due Out Rate (Delayed) = $\frac{\text{Total Line Items Due Out (over 15 days)}}{\text{Total Line Items Authorized}}$

MANAGEMENT ACTIONS THAT CAN BE TAKEN TO REVERSE AN UNFAVORABLE TREND: The whole story is, of course, not told by this indicator. Some bench stock line items are more important than others. Some line items on backorder might be for a quantity of two or three units or a thousand units or more. Some base bench stocks directly support a weapon system where as other bench stocks may support Graphics, Civil Engineers, Transportation, Photo Lab, etc. All bench stocks are important and require constant management surveillance by base supply.

A thorough understanding of how the bench stock system is supposed to work and the procedures for adding and deleting bench stock items is necessary before efforts can be taken to improve an unfavorable "Bench Stock Due Out Rate." Bench stock items are ordered from a variety of sources, i.e., local purchase, GSA, DLA, as well as from military depots. When performing analysis, it is necessary to look at hundreds of variables. A good place to start in resolving problems is by identifying those bench stocks that are contributing the most to the worsening trend. A review of your M-24 Organization Effectiveness Report outlines line items authorized and your supply effectiveness by organization code. A personal visit to the bench stocks being reviewed and discussions with the shop managers is necessary and often enlightening. Likewise, discussions with Base Supply bench stock managers should also be helpful in understanding the problems with a particular bench stock.

Concentrated management efforts can be placed on specific line items that are zero balance. Intensive research and requisition follow-up on each line item will eventually result in an improved bench stock due out rate.

6. OVERALL INVENTORY ACCURACY

IMPORTANCE OF INDICATOR: One of the most important indicators of Chief of Supply performance is Inventory Accuracy. The serviceable balance field on the computer item record is the basis for two computer decision-whether or not to issue assets to satisfy a customer request and when to create a stock replenishment requisition. If the actual warehouse balances differ from item record balances, these decisions will be incorrect. In one case, a warehouse refusal will be generated and customer confidence in the supply support abilities will be impaired. In the other case, the stock replenishment requisition will be delayed and the probability of an ensuing zero balance will increase.

DATA SOURCE: Data is available on the M-32, Monthly Base Supply Management Report.

FORMULA: No external computation for this indicator is necessary, but under program control total units over/short ÷ record balance = % that is minus from 100% = Inventory Accuracy.

MANAGEMENT ACTIONS TO TAKE TO IMPROVE AN UNFAVORABLE INDICATOR TREND: Poor inventory accuracy presents a serious supply situation. Many factors impact on this indicator, i.e., pilferage from warehouse bins or the receiving line, items in the wrong warehouse location, quantity count incorrect by the Receiving Section, incorrect count by the Inventory Section during Cycle or Sample Inventory and inventory counts correct but computer record adjustments incorrect are some of the major problem areas that require review. Sometimes problems with inventory accuracy can be isolated to a single stockroom. This may indicate poor stockroom warehouse management. Also, it could be that the stockroom has gone an extended period without inventory. In some cases, a wall to wall stockroom inventory may be necessary. Some item records may exceed 365 days without inventory, this needs review and correction. Individuals on the night shift may be making issues without issue documents or carelessly pulling stock and returning stock into an incorrect bin. The quality of night work can be checked by requiring posting of issues to a log. The next day, inquiries can be made for stock balance and warehouse locations physically checked. Warehouse security may also require review to be assured that unauthorized personnel do not have access to the warehouse.

The monthly consolidated Inventory Adjustment Document Register M10 categorizes the type of discrepancies identified so that corrective action can be initiated. If the nature of a discrepancy indicates the need for reporting, it will be expeditiously furnished to the OSI Detachment, with information copy to Chief, Security Policy, to determine if there is probable cause to suspect theft and if an investigation is required IAW AFR 125-21. The Inventory section is responsible for obtaining supporting documentation and attaching this documentation to the M10. Completed causative research work sheets or inventory registers containing a statement by the Inventory section that adequate research has been performed and evidence of theft, fraud or misdemeanor (does) (does not) exist.

7. <u>DELINQUENT DOCUMENT RATE</u>

IMPORTANCE OF INDICATOR: The Delinquent Document Rate is a necessary management tool and a good indication of the overall condition of a stock record account. Numerous supply transactions occur each day that require feedback so management can be assured that transactions were completed. A high Delinquent Document Rate normally indicates a poor supply account with unaggressive management. Selected transactions result in the D04 Daily Document Register production of a document control card (DCC). Delinquent Source Documents (DSD) for manually prepared documents such as Post post transactions may experience delays in computer processing. The documents are collected throughout each day and given to the Document Control Section

the same day. They are matched to the output DCC from the D04 the following day. All documents do not become delinquent in the same length of time and all documents do not go into permanent file for an audit trail.

The Document Control Section must be assured that each signed document compares with a DCC card before the document is filed or destroyed. (The documents marked for ultimate destruction are temporarily filed for 15 days, since the Base Service Store (BSS), Tool Issue Center (TIC), Individual Equipment Unit (IEU) went to line item accounting system).

DATA SOURCE: Document control cards are produced by the computer when certain transactions occur. Cards can be used to make a listing for easier reading and use by individuals throughout base supply.

FORMULA: Total documents overdue return to the Document Control Section divided by the total number of Document Control Cards.

MANAGEMENT ACTIONS THAT CAN BE TAKEN TO REVERSE AN UNFAVORABLE TREND: Aggressive management throughout base supply is necessary to hold down the number of delinquent documents. The proper flow of documents must be known by the individuals handling documents and they must give adequate emphasis to the return of signed documents to the Document Control Section. Documents may be missing for a variety of reasons, i.e., property awaiting signature, property not pulled from stock, property not delivered, property issued and the receipt document lost, property lost, property stolen or perhaps missing documents may be accumulating in an office and consequently not delivered to the Document Control Section. A thorough review of all possible causes for the missing document must be accomplished before a certificate of lost document is completed.

8. UNAUTHORIZED EQUIPMENT (ALLOWANCE SOURCE CODE 000 LONGER THAN 15 DAYS)

IMPORTANCE OF INDICATOR: Air Force equipment items are high cost and/or pilferable type items that require accurate control and accountability. The maintenance of accountable records and retaining and maintaining equipment is a costly process. When equipment items are held by custodians without proper authorization an activity elsewhere may be deprived of the item while the unit holding the item is causing unnecessary inventory accounting. In addition, unauthorized items may be taking shop or office space that is needed for another purpose. The impact of an equipment excess within a single organization may not be large but from an entire base or Major Command standpoint the amount of unauthorized equipment might be very significant and require aggressive management attention throughout the command.

DATA SOURCE: Major command programs will normally be established to provide this data. Also, the Q09 Allowance Source Code Listing for ASC 000 provides this information for base review prior to being reported to the data bank by your D16 Daily Equipment Transaction Report or M08 Authorized In Use Report.

FORMULA: This management indicator is usually based on the total dollar value of equipment held without authorization for a period longer than 15 days. Equipment may also be determined by a percentage.

$$\% \text{ of Equipment Excess} = \frac{\#\text{ of units ASC 000 over 15 days}}{\text{TOTAL units authorized}}$$

MANAGEMENT ACTIONS THAT CAN BE TAKEN TO REVERSE AN UNFAVORABLE TREND: In reviewing this indicator it will be helpful to identify the individual items of concern and the custodian accounts on which they were listed. In many cases it is helpful to learn the cause for the excess. There may have been an error in a revised Table of Allowance that incorrectly caused the excess. The item could have been directed into a base by a higher headquarters without establishing appropriate Allowance Source Codes (ASC). Movement of a weapon system from a base may also create excesses. To eliminate the ASC 000 and hold the item, an AF Form 601A with justification must be processed to higher headquarters. This will permit assignment of ASC 000A to the computer record indicating that a request for authorization has been submitted. If retention of the item is not approved, then the excess item must be cleared by turn-in. Occasionally, items will remain on ASC 000 for an extended period because they are extremely large or fastened to the earth or a facility. These kinds of items may require a contract with a civilian business to disconnect or physically move the item. Extensive turn-in delays may occur causing the item to remain as ASC 000 beyond 15 days. Equipment managers should aggressively resolve all ASC 000 beyond 15 days. Problems should be elevated to appropriate commanders when they cannot be resolved at lower management levels.

9. SERVICEABLE BALANCE - NO WAREHOUSE LOCATION

IMPORTANCE OF INDICATOR: Items with a serviceable balance and no warehouse location are lost. Having a serviceable balance without a warehouse location can occur in a number of different ways. Perhaps a receipt would be processed without the item being physically received. Another possibility is that a substitute item may be received and placed in the warehouse location of the preferred item. A receipt could be processed and the item is in the warehouse awaiting location assignment. A receipt could be processed and the item pilfered from the receiving line.

The seriousness of this situation is, of course, that the items are in stock for customer support. If a demand for an item occurs and it cannot be located, a warehouse refusal results. If the customer still requires the item, a requisition through the distribution system will be necessary and delay customer support. Items ordered for War Reserve Materiel (WRM) stock but not assigned a location may degrade the readiness of the WRM kit. This indicator is very important because it informs supply managers about problems with the flow of items without a warehouse location through the receiving and warehouse sections. Closely following this indicator and taking prompt corrective action will be helpful in heading off serious problems and in maintaining an excellent supply account.

DATA SOURCE: M-32 Item Record Data

FORMULA: Serviceable Balance No warehouse location rate = $\frac{\text{Supplies Serviceable no location balance}}{\text{Supplies total item record}}$

MANAGEMENT ACTIONS THAT CAN BE TAKEN TO REVERSE AN UNFAVORABLE TREND: Reversing this unfavorable trend requires quick, aggressive, and positive steps. The problems may be located in the Receiving Section. Perhaps there has been an unusually heavy work load, a new individual assigned to the section or the Receiving Section supervisor has been absent for a period of time. The type of items missing might be a clue that can relate the problem to a particular warehouse stockroom supervisor. Another possibility could be an excessive amount of computer downtime causing "Post Post" document processing. If an extensive amount of incorrect "Post Post" processing has occurred, many management indicators will reflect an unfavorable trend. In any case, early correction of this unfavorable trend must be accomplished.

10. <u>REPAIR CYCLE SUPPORT - DELAYED REPAIRS (EXCLUDING AWP TIME - ALL ORGANIZATIONS)</u>

IMPORTANCE OF INDICATOR: Air Force Repair Cycle assets are high cost repairable type items required in support of a weapon system. These items are extremely limited in our supply system and require close accountability and timely repair. Items not repaired in a timely manner may cause a Not Mission Capable Supply (NMCS) situation to develop. A NMCS condition, as we discussed earlier, causes an additional burden to our distribution system and, of course, causes downtime for a weapon system. This indicator reflects the coordinated effort of supply and maintenance personnel. Supply personnel must assure adequate benchstock for repair and maintain accurate location records. Maintenance must make timely repairs to items and return them serviceable to base supply. This coordination is vital to the mission success of the weapon system.

DATA SOURCE: M-32 Repair Cycle Asset Control Data (all organizations)

FORMULA:

Repair Cycle Support Rate = $\frac{\text{Total units repairable this station over 4 days AFM 66-1 organization \& other organizations}}{\text{Total Units repairable this Station, AFM 66-1 organizations \& total other organizations.}}$

MANAGEMENT ACTIONS THAT CAN BE TAKEN TO REVERSE AN UNFAVORABLE TREND: A daily supply-maintenance meeting is the place to report any problems in the repair cycle system that cannot be resolved at lower working levels. The Chief of Maintenance is extremely interested in the production of his

maintenance shops and the support by base supply. He realizes that excessive delays in repairing repair cycle items may cause an increase work load due to cannibalization. The flying missile or communications mission could also be degraded due to nonavailability of serviceable items. Supply managers need to examine their document control procedures to be assured that adequate controls exist.

SUMMARY

The foregoing is in no way a full treatment of all the management indicators available or used in Base Level Supply Management. The ten indicators selected for review are very important to management but there are other indicators of equal importance. The analysis presented should be helpful to individuals new to the supply "business" in understanding the supply system and becoming an effective contributor to supply management earlier in their supply assignment.

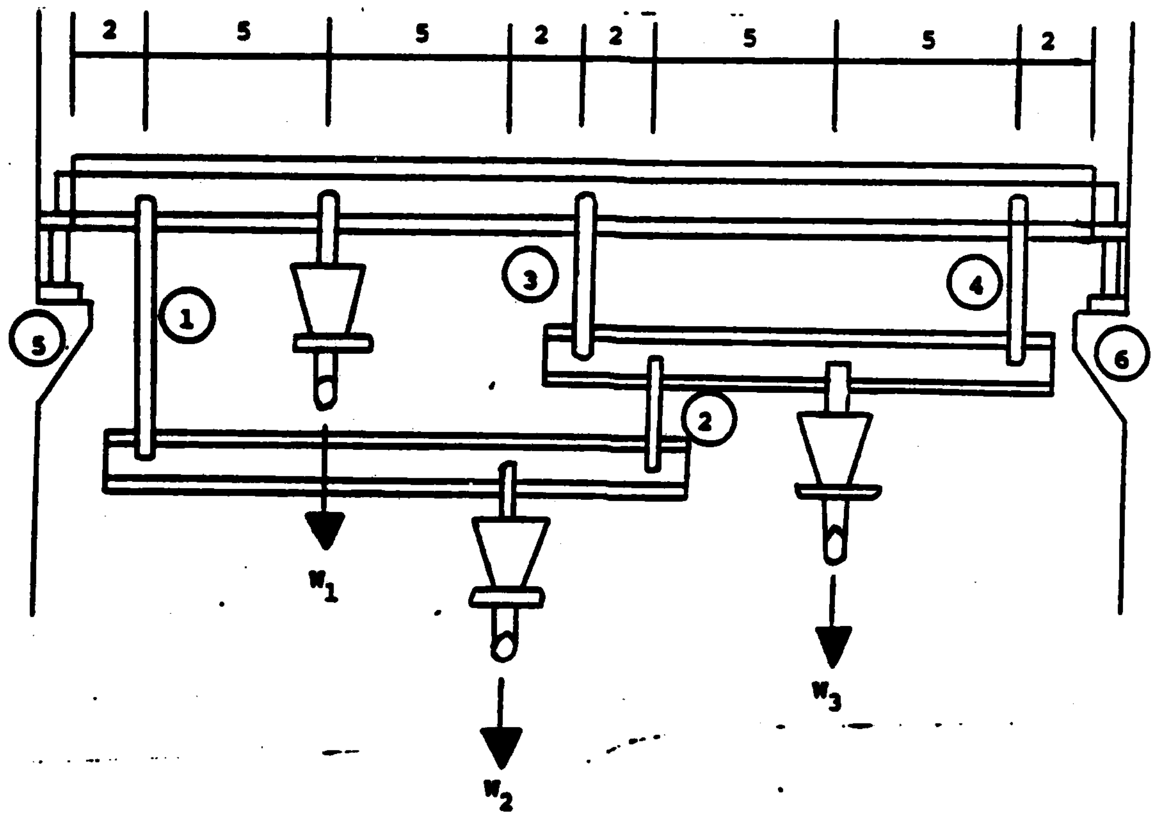
CHAPTER 2

CIVIL ENGINEERING APPLICATIONS

This chapter presents a number of examples designed to illustrate the application of several fundamental management science techniques to typical base civil engineering problems. Three general types of techniques are included: (1) deterministic mathematical modeling (linear programming), (2) network analysis, and (3) engineering economic analysis. The deterministic mathematical programming examples include linear programming, transportation, and assignment problems. The network analysis examples illustrate closed-circuit, shortest route, maximal flow, and PERT/CPM problems. The economic analysis examples illustrate investment decision problems. The solution algorithms applied in these examples are discussed in other chapters.

EXAMPLE 2-1

An overhead crane is being designed for installation in an aircraft maintenance hanger. The crane is configured as follows:



Cables 1 and 2 have an 8-kip capacity. Cables 3 and 4 have a 16-kip capacity. Each crane rail (5 and 6) can support 20 kips. What is the total maximum load that can be carried in its current configuration?

Solution Procedure

A. System Analysis

1. System Components

a. Criterion Variable (Z): Total weight in kips (k); objective is to maximize Z.

b. <u>Decision Variable</u> (W_j): Weight in kips that can be carried by hoist j (j=1,2,and 3).

c. <u>Environmental Variables</u>:

S_i: Load in kips carried by support i (i=1,2,...,5).

2. <u>System Relationships</u>

$Z = W_1 + W_2 + W_3$

S_1: $4W_2 = 14S_1$; $2/7W_2 = S_1 \leq 8k$

S_2: $10W_2 = 14S_2$; $5/7W_2 = S_2 \leq 8k$

S_3: $5W_3 + 10S_2 = 12S_3 \leq 16k$

$5W_3 + 10(5/7W_2) = 12S_3 \leq 16k$

$5/12W_3 + 25/42W_2 = S_3 \leq 16k$

S_4: $2S_2 + 7W_3 = 12S_4 \leq 16k$

$2(5/7W_2) + 7W_3 = 12S_4 \leq 16k$

$5/42W_2 + 7/12W_3 = S_4 \leq 16k$

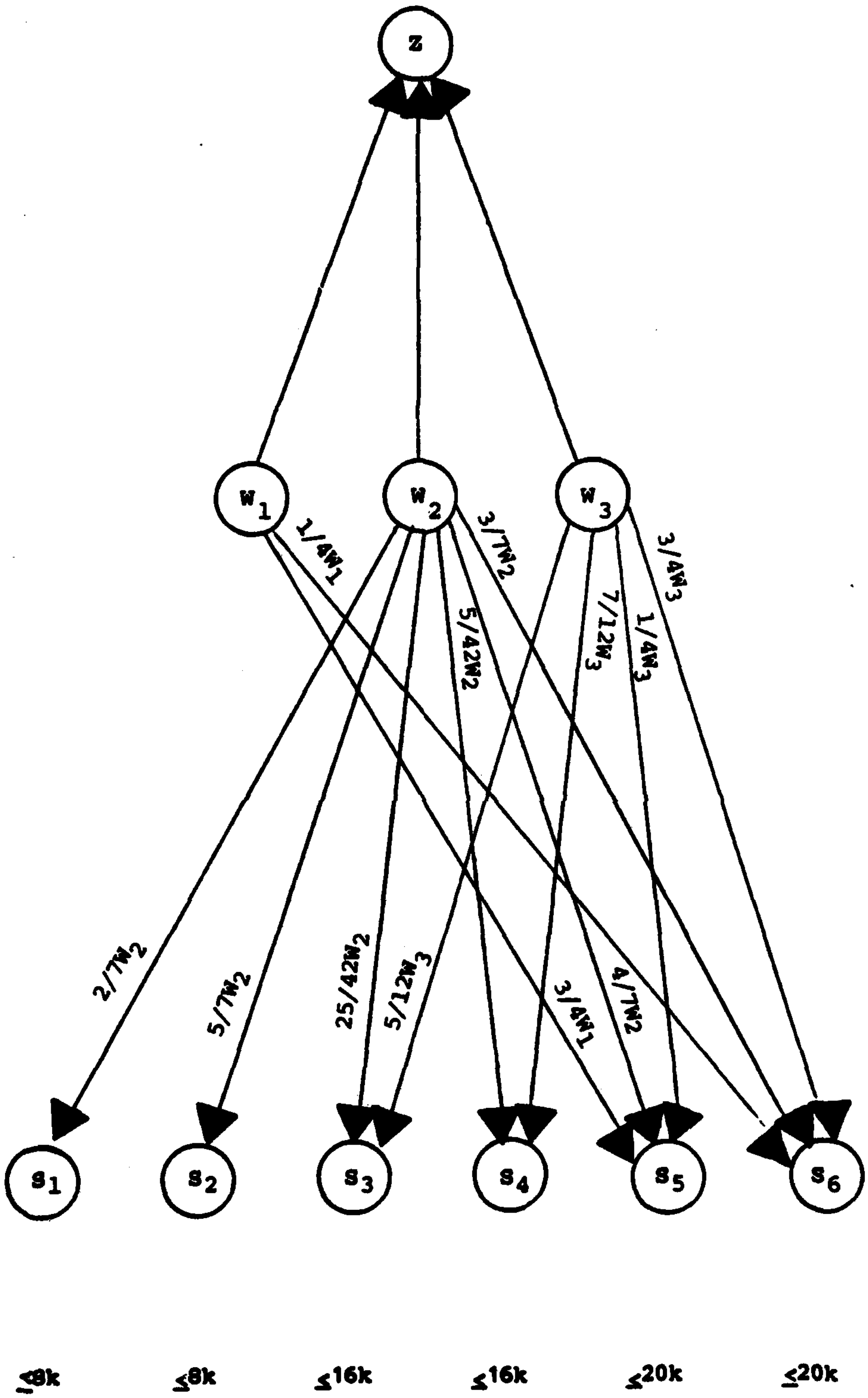
S_5: $7W_3 + 16W_2 + 21W_1 = 28S_5 \leq 20k$

$1/4W_3 + 4/7W_2 + 3/4W_1 \leq 20k$

S_6: $7W_1 + 12W_2 + 21W_3 = 28S_6 \leq 20k$

$1/4W_1 + 3/7W_2 + 3/4W_3 = S_6 \leq 20k$

Figure 2-1


Z
W1
W2
W3
1/4W1
3/7W2
3/4W3
5/42W2
7/12W3
1/4W3
2/7W2
5/7W2
25/42W2
5/12W3
3/4W1
4/7W2
S1
S2
S3
S4
S5
S6
≤8k
≤8k
≤16k
≤16k
≤20k
≤20k

B. **LP Model**

Maximize: $Z = W_1 + W_2 + W_3$ (0)

$$
\begin{array}{llcl}
\text{S.T.} & 2/7W_2 & \leq 8 & (1) \\
& 5/7W_2 & \leq 8 & (2) \\
& 23/42W_2 + 5/12W_3 & \leq 16 & (3) \\
& 5/42W_2 + 7/12W_3 & \leq 16 & (4) \\
& 3/4W_1 + 4/7W_2 + 1/4W_3 & \leq 20 & (5) \\
& 1/4W_1 + 3/7W_2 + 3/4W_3 & \leq 20 & (6) \\
& W_1, W_2, W_3, & \geq 0 &
\end{array}
$$

C. **Solution Algorithm**

Table 2-1

	z	W_1	W_2	W_3	S_1	S_2	S_3	S_4	S_5	S_6	bi	ri
	1	-1	-1	-1	0	0	0	0	0	0	0	
S_1			2/7		1						8	
S_2			5/7			1					8	
S_3			25/42	5/12			1				16	
S_4			5/42	7/12				1			16	
S_5		3/4	4/7	1/4					1		20	26 2/3
S_6		1/4	3/7	3/4						1	20	80

The initial solution is not optimal. There is a tie for the entering variable. Arbitrarily select W_1. As W_1 enters, S_5 leaves. The new solution is:

Table 2-2

	Z	W_1	W_2	W_3	S_1	S_2	S_3	S_4	S_5	S_6	bi	ri
	1	0	-5/21	-2/3					4/3		80/3	
S_1			2/7		1						8	
S_2			5/7			1					8	
S_3			25/42	5/12			1				16	38.4
S_4			5/42	7/12				1			16	27.4
W_1		1	16/21	1/3					4/3		80/3	80
S_6		0	5/21	2/3					-1/3	1	40/3	20

Solution is not optimal. W_3 enters; S_6 leaves. The new solution is:

Table 2-3

	Z	W_1	W_2	W_3	S_1	S_2	S_3	S_4	S_5	S_6	bi
	1	0	0	0					1	1	40
S_1			2/7		1						8
S_2			5/7			1					8
S_3		0	25/56	0			1		5/24	-5/8	46/6
S_4		0	-5/56	0				1	7/24	-7/8	26/6
W_1		1	9/14	0					3/2	-1/2	20
W_3		0	5/14	1					-1/2	3/2	20

Solution is optimal. However, there are alternative optimal solutions, i.e., W_2 can be brought into the basis without changing Z. If this is done, the solution is:

W_1 = 64/6 Z = 40

W_2 = 56/5

W_3 = 16

EXAMPLE 2-2

A contractor will submit a unit-price bid on the foundation work for a large concrete structure to be constructed in support of development work on the MX missile system. After reviewing the plans and specifications and his estimate of the competition on this project, he sets a total bid limit at \$750,000 (7.5×10^5). The government's estimate and the contractor's estimate of the quantities of work involved are shown in the table below. The last column of the table is the contractor's estimate of the acceptable range within which he can submit unit prices. Given the estimated construction schedule in the table and an interest rate of 10%, determine unit prices for the three required construction items which will maximize the present worth of his income.

ITEM	QTY. EST		CONSTR. SCHEDULE			UNIT PRICE RANGE
	GOV'T.	CONTR.	YR 1	YR 2	YR 3	
Earth Excavation	4×10^5	3.5×10^5	2.5×10^5	1.0×10^5		\$0.80-1.80/CY
Rock Excavation	3×10^4	3.5×10^4	0.5×10^4	3.0×10^4		\$6.00-11.00/CY
Drilling	1.8×10^3	1.8×10^3		0.5×10^3	1.3×10^3	\$6.00-12.00/ft

Solution Procedure

A. System Analysis

1. System Component

a. Criterion Variable (Z): Present value in dollars of total income; the objective is to maximize Z.

b. Decision Variables (X_i): Unit price in dollars for task i (i=1,2,3).

c. Environmental Variables:

(1) UL_i = upper limit on unit price for i

(2) LL_i = lower limit on unit price for i

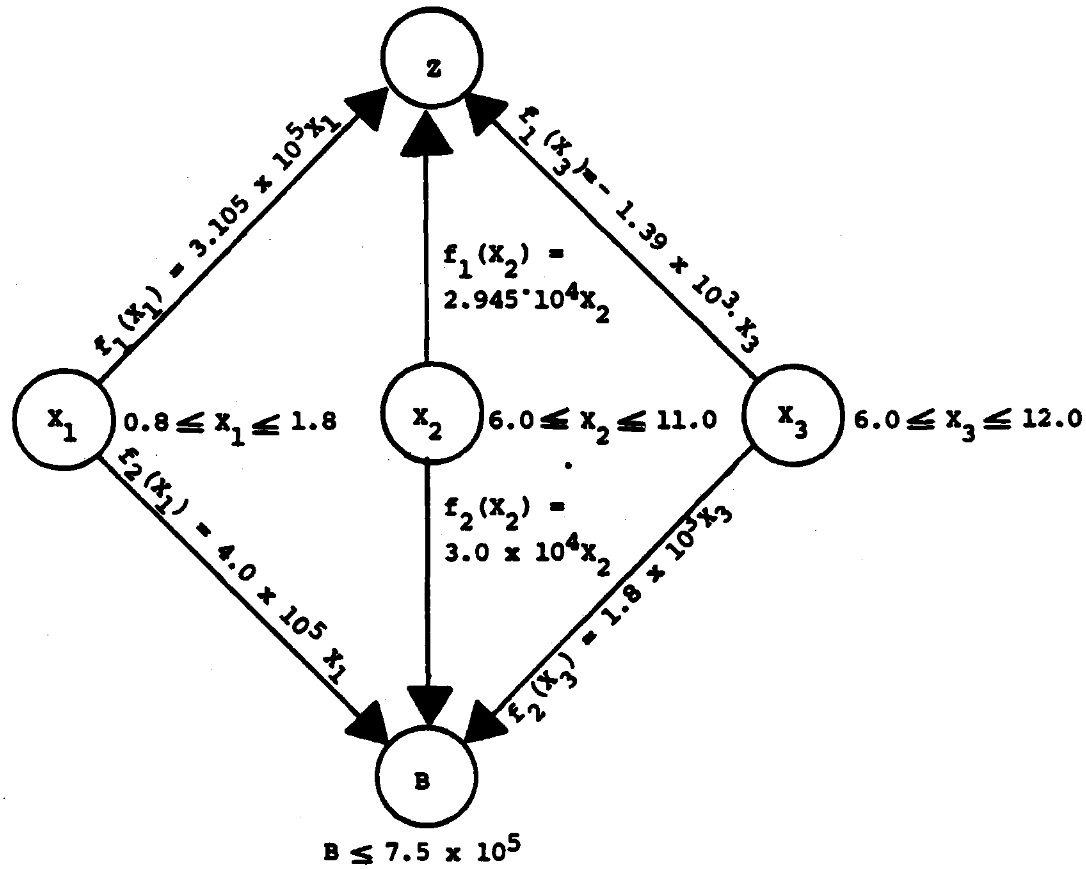
(3) B = max. allowable bid price

(4) I = interest rate

(5) X_{ij} = amount of item i scheduled for year j (i=1,2,3; j=1,2,3)

2. **System Relationships**

Figure 2-2



where:

$$f_1(X_1) = \left[\frac{1}{(1.10)^1} \quad (2.5 \times 10^5) + \frac{1}{(1.10)^2} \quad (1.0 \times 10^5)\right] X_1$$

$$= \left[(0.91)(2.5 \times 10^5) + (0.83)(1.0 \times 10^5)\right] X_1$$

$$= (2.275 \times 10^5 + 0.83 \times 10^5) X_1$$

$$= 3.105 \times 10^5 X_1$$

$$f_1(X_2) = \left[\frac{1}{(1.10)^1} \quad (0.5 \times 10^4) + \frac{1}{(1.10)^2} \quad (3.0 \times 10^4)\right] X_2$$

$$= \left[(0.91)(0.5 \times 10^4) + (0.83)(3.0 \times 10^4)\right] X_2$$

$$f_1(X_2) = (0.455 \times 10^4 + 2.49 \times 10^4) X_2$$

$$= 2.945 \times 10^4 \cdot X_2$$

$$f_1(X_3) = \left[\frac{1}{(1.10)^2}\ (0.5\times10^3) + \frac{1}{(1.10)^3}\ (1.3\times10^3)\right] X_3$$

$$= \left[(0.83)(0.5\times10^3) + (0.75)(1.3\times10^3)\right] X_3$$

$$= (0.415\times10^3 + 0.975\times10^3) X_3$$

$$= 1.390\times10^3 \cdot X_3$$

$$f_2(X_1) = 4.0\times10^5 X_1$$

$$f_2(X_2) = 3.0\times10^4 X_2$$

$$f_2(X_3) = 1.8\times10^3 X_3$$

B. <u>LP Model</u>

Maximize: $Z=3.105\times10^5 X_1+2.945\times10^4 X_2+1.39\times10^3 \cdot X_3$ (0)

S.T. $4.0\times10^5 X_1+3.0\times10^4 X_2+1.8\times10^3 X_3 \leq 7.5\times10^5$ (1)

$X_1 \geq 0.80$ (2)

$X_1 \leq 1.80$ (3)

$X_2 \geq 6.00$ (4)

$X_2 \leq 11.00$ (5)

$X_3 \geq 6.00$ (6)

$X_3 \leq 12.00$ (7)

C. <u>Solution Algorithm</u>

Max.: $Z=310,500X_1 + 29,450X_2 + 1,390X_3 + 0S_1 + 0E_2 - MA_2 + 0S_3 + 0E_4 - MA_4 + 0S_5 + 0E_6 - MA_6 + 0S_7$ (0)

S.T. $400,000X_1 + 30,000X_2 + 1,800X_3 + S_1 = 750,000$ (1)

$X_1 - E_2 + A_2 = 0.80$ (2)

$X_1 + S_3 = 1.80$ (3)

$X_2 - E_4 + A_4 = 6.00$ (4)

$X_2 + S_5 = 11.00$ (5)

$X_3 - E_6 + A_6 = 6.00$ (6)

$X_3 + S_7 = 12.00$ (7)

Table 2-5

I

BASIS	Z	X_1	X_2	X_3	S_1	E_2	A_2	S_3	E_4	A_4	S_5	E_6	A_6	S_7	b_i	r_i
	1	-310,500	-29,450	-1,390	0	0	+M	0	0	+M	0	0	+M	0	0	
		400,000	30,000	1,800	1										750,000	
		1				-1	1								0.8	
		1						1							1.8	
			1						-1	1					6.0	
			1								1				11.0	
				1								-1	1		6.0	
				1										1	12.0	

Modifying to obtain an initial feasible solution:

Table 2-6

II

BASIS	Z	X_1	X_2	X_3	S_1	E_2	A_2	S_3	E_4	A_4	S_5	E_6	A_6	S_7	b_i	r_i
	1	-310,500 -M	-29,450 -M	-1,390 -M	0	M	0	0	M	0	0	M	0	0	-12.8M	
S_1		400,000	30,000	1,800	1										750,000	1.875
A_2		1				-1	1								0.8	0.8
S_3		1						1							1.8	1.8
A_4			1						-1	1					6.0	
S_5			1								1				11.0	
A_6				1								-1	1		6.0	
S_7				1										1	12.0	

Solution is not optimal.

X_1 enters

A_2 leaves

Table 2-7

III

	Z	X_1	X_2	X_3	S_1	E_2	A_2	S_3	E_4	A_4	S_5	E_6	A_6	S_7	b_i	r_i
	1	0	-29,450 -M	-1,390 -M	0	-310,000	310,500 +M	0	M	0	0	M	0	0	248,400 -12M	
S_1		0	30,000	1,800	1	400,000	-400,000								430,000	14.33
X_1		1				-1	1								0.8	
S_3		0				1	-1	1							1.0	
A_4			1						-1	1					6.0	6.0
S_5			1								1				11.0	11.0
A_6				1								-1	1		6.0	
S_7				1										1	12.0	

Solution is not optimal. X_2 enters, A_4 leaves.

Table 2-8

IV

BASIS	Z	X_1	X_2	X_3	S_1	E_2	A_2	S_3	E_4	A_4	S_5	E_6	A_6	S_7	b_i	r_i
	1	0	0	-1,390 -M	0	-310,500	310,500 +M	0	-29,450	29,450 +M	0	M	0	0	425,100 -6M	
S_1		0	0	1,800	1	400,000	-400,000		30,000	-30,000					250,000	138.8
X_1		1				-1	1								0.8	
S_3		0				1	-1	1							1.0	
X_2			1						-1	1					6.0	
S_5		0	0						1	-1	1				5.0	
A_6				1								-1	1		6.0	6.0
S_7				1										1	12.0	12.0

Solution is not optimal. X_3 enters, A_6 leaves.

V

Table 2-9

BASIS	Z	X_1	X_2	X_3	S_1	E_2	A_2	S_3	E_4	A_4	S_5	E_6	A_6	S_7	b_i	r_i
	1	0	0	0	0	-310,500	310,500 +M	0	-29,450	29,450 +M	0	-1390	1390 +M	0	433,440	
S_1		0	0	0	1	400,000	-400,000		30,000	-30,000		1800	-1800		239,200	.598
X_1		1				-1	1								0.8	-
S_3		0				1	-1	1							1.0	1.0
X_2			1						-1	1					6.0	
S_5									1	-1	1				5.0	
X_3				1								-1	1		6.0	
S_7				0								1	-1	1	6.0	

Solution is not optimal. E_2 enters, S_1 leaves.

VI

Table 2-10

	Z	X_1	X_2	X_3	S_1	E_2	A_2	S_3	E_4	A_4	S_5	E_6	A_6	S_7	b_i	r_i
	1	0	0	0	0	0	M	0	-6163	6163	0	7	M-7	0	619,119	
E_2					0	1	-1		.075	-.075		.0045	-.0045		.598	7.97
X_1		1				0	0		.075	-.075		.0045	-.0045		1.398	18.64
S_3						0	0	1	.075	-.075		.0045	-.0045		1.598	21.31
X_2			1						-1	1					6.0	-
S_5									1	-1	1				5.0	5.0
X_3				1								-1	1		6.0	-
S_7												1	-1	1	6.0	-

Solution is not optimal. E_4 enters, S_5 leaves

Table 2-11

	Z	X_1	X_2	X_3	S_1	E_2	A_2	S_3	E_4	A_4	S_5	E_6	A_6	S_7	b_i	r_i
	1	0	0	0	0	0	M	0	0	0	6163	7	M-7	0	649,934	
E_2						1	-1		0	0	-.075	.0045	-.0045		0.973	
X_1		1							0	0	-.075	.0045	-.0045		1.773	
S_3						0	0	1	0	0	-.075	.0045	-.0045		1.973	
X_2			1						0	0	1				11.0	
E_4									1	-1	1				5.0	
X_3				1								-1	1		6.0	
S_7												1	-1	1	6.0	

Solution is optimal.

EXAMPLE 2-3

The base civil engineer has established three locations around the base at which sand and salt stockpiles are maintained for winter icing and snow storms. For most storms, all salt and sand is distributed from these three locations to four major zones around the base. The Chief of Operations and Maintenance wants to determine the best way to allocate the materials available at each stockpile to the various service zones. He is primarily concerned with the time required to meet the needs of the four zones to be serviced. While travel times are not available, the following table summarizes the average distance from the various supply points to the four service zones. The table also indicates the number of truckloads of material available at each storage location and the estimated number of truckloads required at each service zone:

Table 2-12

STORAGE SITE	SERVICE ZONE 1	2	3	4	TRUCKLOADS AVAILABLE
1	1.2	3.4	2.2	2.5	200
2	3.8	4.5		2.3	250
3	4.3	6.4	3.5	4.2	150
Trucks Req'd	125	150	125	150	600 / 550

It is not feasible to deliver any material from storage site 2 to service zone 3. How should the demand at each of the four service zones be satisfied to minimize the total delivery time?

Solution Procedure

A. **System Analysis**

1. **System Components**

a. **Criterion Variable** (Z): Total truckload-miles; the objective is to minimize Z.

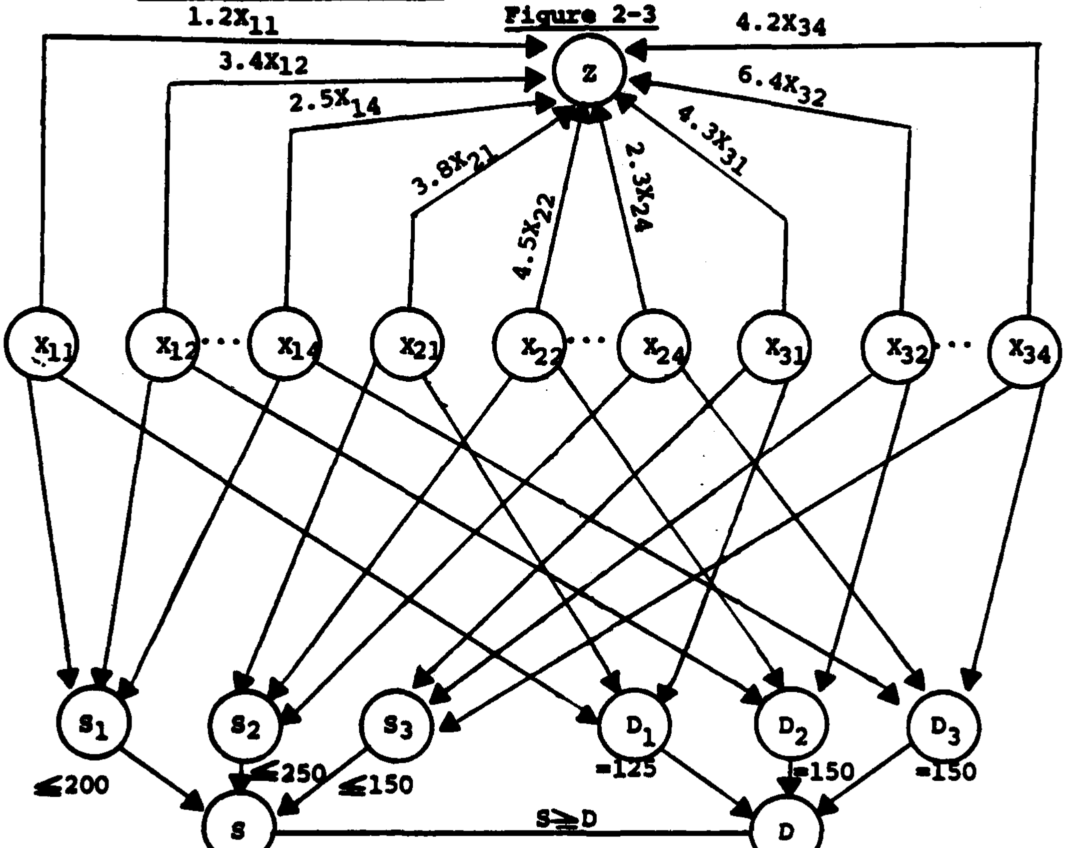
b. **Decision Variable** (X_{ij}): Truckloads of material i to service zone j, where (i=1,2,3; j=1,2,3,4).

c. **Environmental Variables**:

S_i = truckloads of material available at storage site i (i=1,2,3).
D_j = truckloads of material required at service zone j (j=1,2,3,4)
S = total supply at all sites
D = total requirement at all zones.

2. **System Relationships**

Figure 2-3



B. <u>LP Model</u>

$$\text{Minimize} \quad Z = 1.2X_{11} + 3.4X_{12} + 2.2X_{13} + 2.5X_{14} + 3.8X_{21} + 4.5X_{22} + 2.3X_{24} + 4.3X_{31} + 6.4X_{32} + 3.5X_{33} + 4.2X_{34}$$

$$\begin{aligned}
\text{S.T.} \quad & X_{11} + X_{12} + X_{13} + X_{14} \leq 200 \\
& X_{21} + X_{22} + X_{24} \leq 250 \\
& X_{31} + X_{32} + X_{33} + X_{34} \leq 150 \\
& X_{11} + X_{21} + X_{31} = 125 \\
& X_{12} + X_{22} + X_{32} = 150 \\
& X_{13} + X_{33} = 125 \\
& X_{14} + X_{24} + X_{34} = 150 \\
& X_{ij} \geq 0 \quad (i=1,2,3;\ j=1,2,3,4)
\end{aligned}$$

C. <u>Solution Algorithm</u>

This LP model can be solved with the general simplex algorithm. However, this problem can be solved somewhat more efficiently using the transportation (simplex) algorithm:

Table 2-13

i \ j	1	2	3	4	5(D)	S_i
1	1.2 125	3.4 75	2.2	2.5	0	200
2	3.8	4.5 75	M 125	2.3 50	0	250
3	4.3	6.4	3.5	4.2 100	0 50	150
D_j	125	150	125	150	50	600

(1,3) $+2.2 - M + 4.5 - 3.4 = -M+3.3$
(1,4) $+2.5 - 2.3 + 4.5 - 3.4 = 1.3$
(1,5) $0 - 0 + 4.2 - 2.3 + 4.5 - 3.4 = 3.0$
(2,1) $3.8 - 1.2 + 3.4 - 4.5 = 1.5$
(2,5) $0 - 0 + 4.2 - 2.3 = 1.9$
(3,1) $4.3 - 1.2 + 3.4 - 4.5 + 2.3 - 4.2 = 0.1$
(3,2) $6.4 - 4.5 + 2.3 - 4.2 = 0$
(3,3) $3.5 - M + 2.3 - 4.2 = -M+1.6$

Solution is not optimal; (3,3) enters, X_{34} leaves; X_{34}=100, X_{24}=150, X_{23}=25.

Table 2-14

i \ j	1	2	3	4	5(D)	S_i
1	1.2 125	3.4 75	2.2	2.5	0	200
2	3.8	4.5 75	M 25	2.3 150	0	250
3	4.3	6.4	3.5 100	4.2	0 50	150
D_j	125	150	125	150	50	600

(1,3) 2.2-M+4.5-3.4 = -M+3.3

(1,4) 2.5-2.3+4.5-3.4 = 1.3

(1,5) 0 - 0 + 3.5 - M + 4.5 - 3.4 = -M+4.6

(2,1) 3.8 - 1.2 + 3.4 - 4.5 = 1.5

(2,5) 0 - 0 + 3.5 - M = -M+3.5

(3,1) 4.3 - 1.2 + 3.4 - 4.5 + M - 3.5 = M-1.5

(3,2) 6.4 - 4.5 - M + 3.5 = -M+5.4

(3,4) 4.2 - 3.5 + M - 2.3 = M-1.6

Solution is not optimal; (1,3) enters; X_{34} leaves and X_{12} = 50 X_{13} = 25
X_{22} = 100

Table 2-15

i \ j	1	2	3	4	5(D)	S_i
1	1.2 125	3.4 50	2.2 25	2.5	0	200
2	3.8	4.5 100	M	2.3 150	0	250
3	4.3	6.4	3.5 100	4.2	0 50	150
D_j	125	150	125	150	50	600

(1,4) $2.5 - 2.3 + 4.5 - 3.4 = 1.3$
(1,5) $0 - 0 + 3.5 - 2.2 = 1.3$
(2,1) $3.8 - 1.2 + 3.4 - 4.5 = 1.5$
(2,3) $M - 2.2 + 3.4 - 4.5 = M-3.3$
(2,5) $0 - 0 + 3.5 - 2.2 + 3.4 - 4.5 = 0.2$
(3,1) $4.3 - 1.2 + 2.2 - 3.5 = 1.8$
(3,2) $6.4 - 3.4 + 2.2 - 3.5 = 1.7$
(3,4) $4.2 - 3.5 + 2.2 - 3.4 + 4.5 - 2.3 = 1.7$

Solution is optimal:

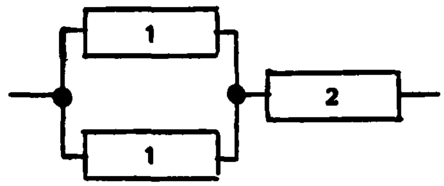
$X_{11} = 125 \times 1.2 = 150.0$
$X_{12} = 50 \times 3.4 = 170.0$
$X_{13} = 25 \times 2.2 = 55.0$
$X_{22} = 100 \times 4.5 = 450.0$
$X_{24} = 150 \times 2.3 = 345.0$
$X_{33} = 100 \times 3.5 = 350.0$
$X_{35} \quad 50 \times 0 \quad 0$

1520 truckload-miles.

EXAMPLE 2-4

A particular subsystem is extensively used in many of the mechanical systems on base. There are 1,000 of these subsystems in use, each of which has two components. Each subsystem assembly is configured as follows:

Figure 2-4



Component 1 has a reliability of 0.925 and component 2 has a reliability of 0.90. Replacement components are available from two sources. They can be purchased from a vendor in the year in which a failure occurs. Alternatively, half of the components that fail in one year can be repaired and reconditioned for use the next year. Purchase and repair costs for each of these components are shown below:

Table 2-16

	COMPONENT	
	1	2
PURCHASE	25	50
REPAIR	20	35

Considering only the next two years, what is the present value of money to be programmed for component purchase and component repair (assume an interest rate of 10%).

Solution Procedure

A. System Analysis

1. Components

a. Criterion Variable (Z): Present value of the total replacement cost; objective to minimize Z.

b. Decision Variable (X_{ij}): Number of units of type i for year j where:

i=1: purchased units of component 1
  2: purchased units of component 2
  3: repaired units of component 1
  4: repaired units of component 2

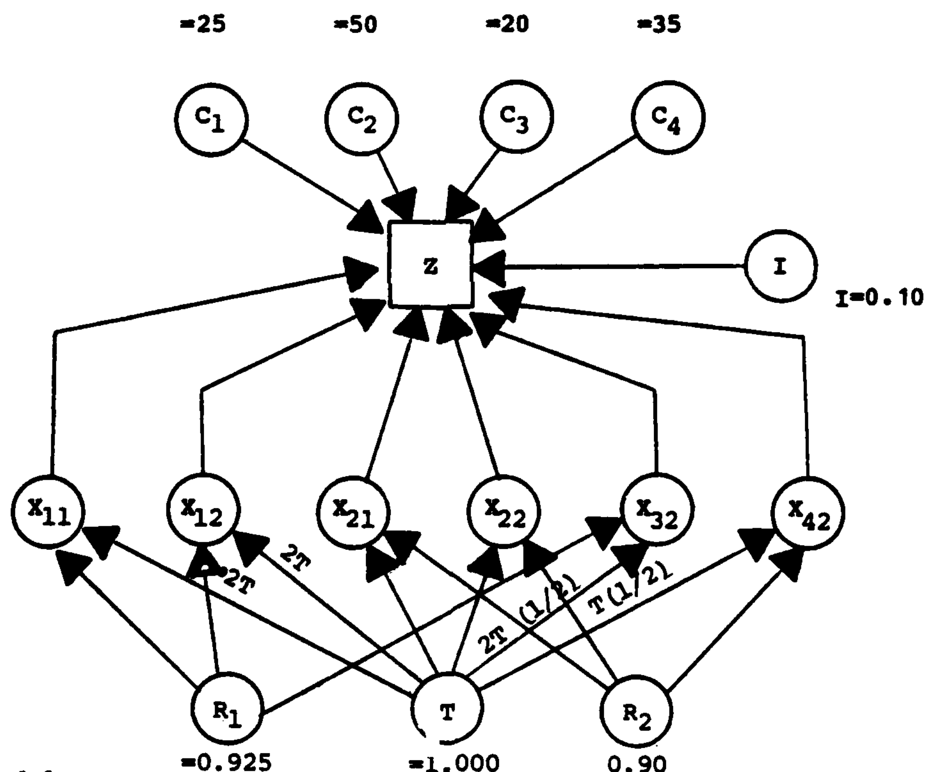
j=1,2

c. <u>Environmental Variables</u>

R_1: reliability of component 1
R_2: reliability of component 2
T : total number of subsystems
I : interest rate
C_i: cost of compound i

2. <u>System Relationships</u>

<u>Figure 2-5</u>



B. <u>LP Model</u>

Minimize: $$Z = \left[\frac{1}{(1.10)^1} \cdot 25 \cdot X_{11} + \frac{1}{(1.10)^2} \cdot 25 \cdot X_{12} + \frac{1}{(1.10)^1} \cdot 50 \cdot X_{21} + \frac{1}{(1.10)^2} \cdot 50 \cdot X_{22} + \frac{1}{(1.10)^2} \cdot 20 \cdot X_{32} + \frac{1}{(1.10)^2} \cdot 35 \cdot X_{42} \right]$$

$$
\begin{aligned}
\text{S.T.}\quad & X_{11} && = 150 \\
& X_{21} && = 100 \\
& X_{12} + X_{32} && = 150 \\
& X_{22} + X_{42} && = 100 \\
& X_{11} + X_{12} && \leq 225 \\
& X_{21} + X_{22} && \leq 150 \\
& X_{32} && \leq 75 \\
& X_{42} && \leq 50 \\
& X_{ij} \geq 0 \ (i=1,2,\ldots,4;\ j=1,2)
\end{aligned}
$$

C. Solution Algorithm

The problem can be modeled as a transportation system with the following sources and demand points:

Source: 1 = purchased units of component 1
(i) 2 = purchased units of component 2
3 = repaired units of component 1
4 = repaired units of component 2

Demand: 1 = component 1 required in year 1
(j) 2 = component 2 required in year 1
3 = component 1 required in year 2
4 = component 2 required in year 2

Table 2-17

i \ j	1	2	3	4	S_i
1	22.7 150	M ε	20.7 75	M	225
2	M	45.5 100	M	41.3 50	150
3	M	M	16.5 75	M	75
4	M	M	M	28.9 50	50
D_j	150	100	150	100	500

Initial solution by least cost method is degenerate; arbitrarily place ε in (1,2).

Checking for optimality:

(1,4): $M - 41.3 + 45.5 - \varepsilon = M+4.2$
(2,1): $M - 22.7 + M - 45.5 = 2M-68.2$
(2,3): $M - 45.5 + M - 20.7 = 2M-66.2$
(3.1): $M - 22.7 + 20.7 - 16.5 = M-18.5$
(3,2): $M - M + 20.7 - 16.5 = 4.2$
(3,4): $M - 41.3 + 45.5 - M + 20.7 - 16.5 = 8.4$
(4,1): $M - 22.7 + M - 45.5 + 41.3 - 28.9 = 2M-55.8$
(4,2): $M - 45.5 + 41.3 - 28.9 = M-33.1$
(4,3): $M - 20.7 + M - 45.5 + 41.3 - 28.9 = M-53.8$

Solution is optimal.

$$
\begin{aligned}
Z &= 150(22.7) + 75(20.7) + 100(45.5) + 50(41.3) + 75(16.5) + 50(28.9) \\
&= 3405 + 1552.5 + 4550 + 2065 + 1237.5 + 1445 \\
&= 14{,}255
\end{aligned}
$$

<u>EXAMPLE 2-5</u>

Four jobs are to be completed by four structural maintenance and repair crews. The superintendent and foreman have estimated the time required for each crew to complete each job:

Table 2-18

	j \ i	TASK 1	2	3	4
CREW	1	10	25	16	11
	2	7	26	13	21
	3	35	19	18	16
	4	19	26	24	10

How should the crew-task assignments be made to minimize the total time required to complete these tasks?

<u>Solution Procedure</u>

A. <u>System Analysis</u>

1. <u>System Components</u>

a. <u>Criterion Variable</u> (Z): Total time to complete all task;

objective is to minimize Z.

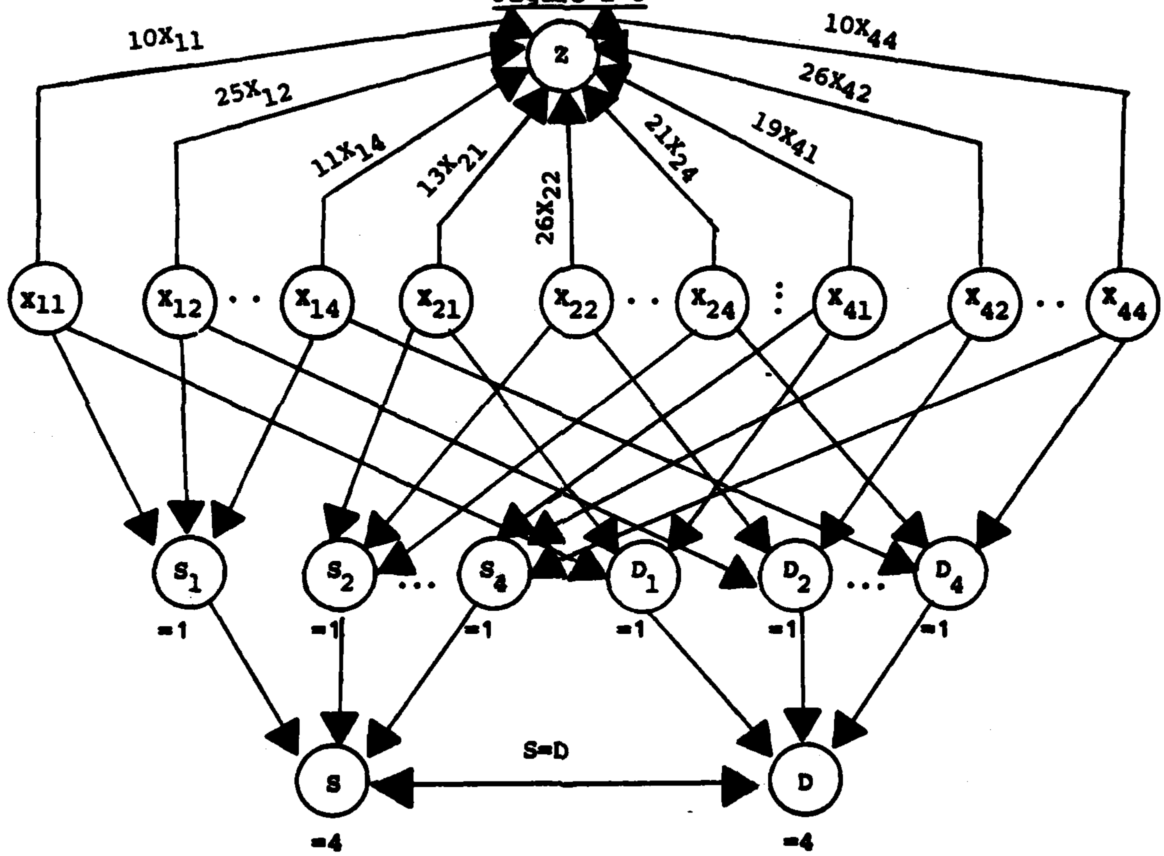
b. <u>Decision Variable</u> (X_{ij}): X_{ij} = 1 if crew i is assigned to task j (i,j = 1,2,...,4); 0 otherwise.

c. <u>Environmental Variable</u>:

S_i = no. of tasks assigned to crew i (i=1,2,...,4) S_i=1
D_j = no. of crews assigned to task j (j=1,2,...,4) D_j=1
S = total no. of crews (4)
D = total no. of tasks (4)

2. <u>System Relationships</u>

<u>Figure 2-6</u>



B. <u>LP Model</u>

Minimize: $Z = 10X_{11} + 25X_{12} + 16X_{13} + 11X_{14} + 13X_{21} + 26X_{22} + 7X_{23} + 21X_{24} + 35X_{31} + 19X_{32} + 18X_{33} + 16X_{34} + 19X_{41} + 26X_{42} + 24X_{43} + 10X_{44}$

S.T. $X_{11} + X_{12} + X_{13} + X_{14} = 1$

$X_{21} + X_{22} + X_{23} + X_{24} = 1$

$X_{31} + X_{32} + X_{33} + X_{34} = 1$

$X_{41} + X_{42} + X_{43} + X_{44} = 1$

$X_{11} + X_{21} + X_{31} + X_{41} = 1$

$X_{12} + X_{22} + X_{32} + X_{42} = 1$

$X_{13} + X_{23} + X_{33} + X_{43} = 1$

$X_{14} + X_{24} + X_{34} + X_{44} = 1$

$X_{ij} = 0,1 \ (i,j=1,2,3,4)$

C. <u>Solution Algorithm</u>

This is a basic Assignment Model. Using the Hungarian algorithm:

Table 2-19

i \ j	1	2	3	4	
1	0	15	6	1	10
2	0	19	6	14	7
3	19	3	2	0	16
4	9	16	14	0	10
					43

Table 2-20

i \ j	1	2	3	4	
1	0	12	4	1	
2	0	16	4	14	
3	19	0	0	0	
4	9	13	12	0	
		3	2		5
					48

Solution is not optimal. The new matrix is:

Table 2-21

i \ j	1	2	3	4
1	0	11	3	0
2	0	15	3	13
3	20	0	0	0
4	10	13	12	0

Solution is not optimal. The new matrix is:

Table 2-22

i \ j	1	2	3	4
1	$\boxed{0}$	8	0	0
2	0	12	$\boxed{0}$	13
3	23	$\boxed{0}$	0	3
4	10	10	9	$\boxed{0}$

Optimal assignment can be made:

$$X_{11}, X_{23}, X_{32}, X_{44} = 1$$

$$Z = 10 + 13 + 19 + 10 = \underline{52}$$

<u>EXAMPLE 2-6</u>

Five different types of dewatering pumps are available for use at five different construction sites. The efficiency of each pump in producing maximum yield at each well is shown in this table:

Table 2-23

		SITE				
		1	2	3	4	5
PUMP TYPE	1	50	30	70	40	50
	2	60	40	60	30	60
	3	30	60	50	60	30
	4	60	80	40	70	80
	5	20	50	30	80	60

Determine the pump type-site assignment that will maximize total efficiency for all sites.

Solution Procedure

A. System Analysis

1. System Components

a. Criterion Variable (Z): Total pump efficiency; the objective is to maximize Z.

b. Decision Variables (X_{ij}): X_{ij} = 1: if pump type i is assigned to site j (i=j=1,2,...,5) =0 otherwise.

c. Environmental Variables

S_i = number of sites to which pump type i can be assigned (S_i=1)

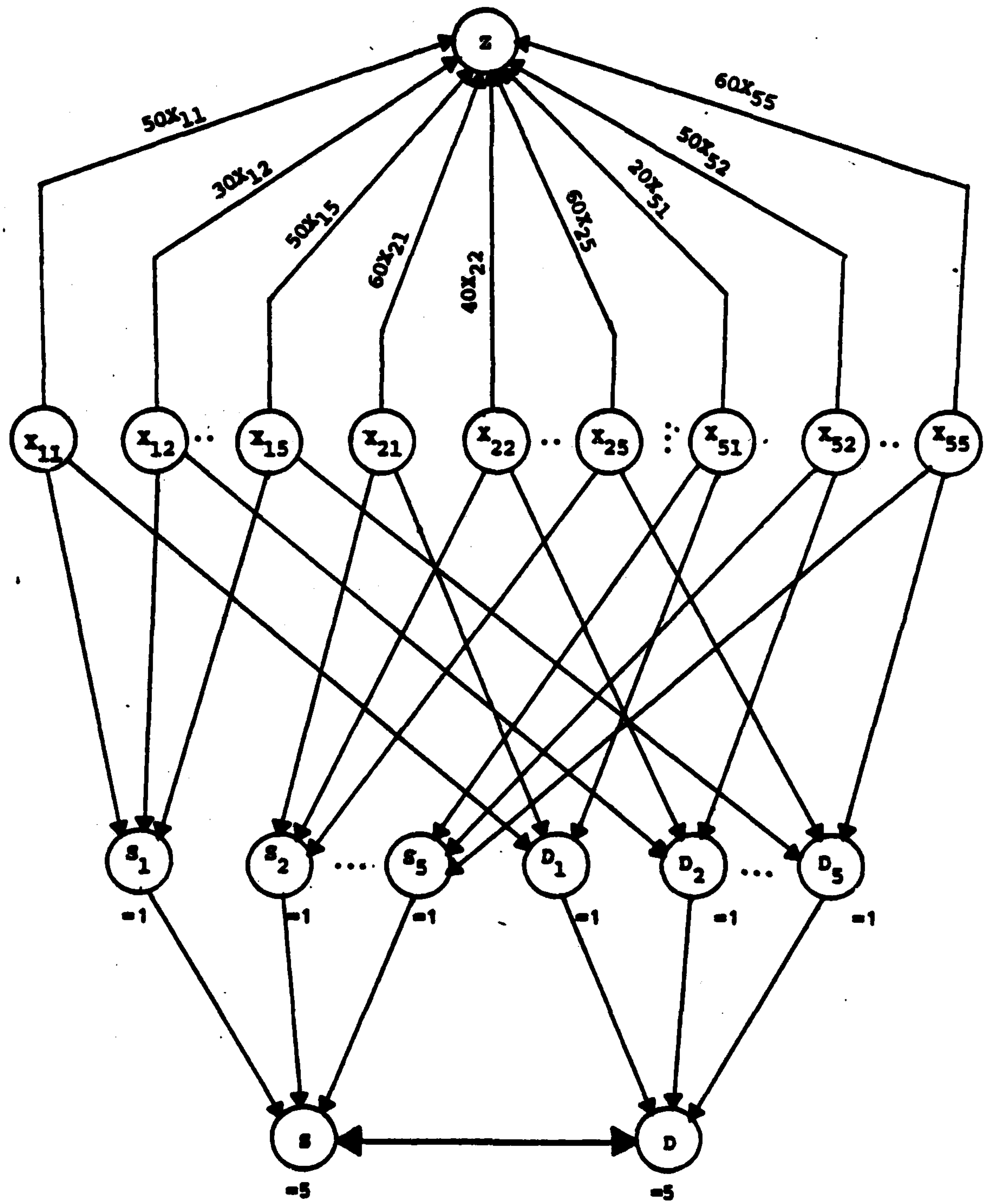
D_j = number of pump types which can be assigned to site j (D_j=1)

S = total number of pump types

D = total number of sites

2. **Systems Relationships**

Figure 2-7



$X_{ij} = 0,1$

B. **<u>LP Model</u>**

$$
\begin{aligned}
\textbf{Maximize:}\quad Z \;=\;& 50X_{11} + 30X_{12} + 70X_{13} + 40X_{14} + 50X_{15} + 60X_{21} + 40X_{22} \\
&+ 60X_{23} + 30X_{24} + 60X_{25} + 30X_{31} + 60X_{32} + 50Y_{33} + 60X_{34} \\
&+ 30X_{35} + 60X_{41} + 80X_{42} + 40X_{43} + 70X_{44} + 80X_{45} + 20X_{51} \\
&+ 50X_{52} + 30X_{53} + 80X_{54} + 60X_{55}
\end{aligned}
$$

$$\text{S.T.}\qquad \sum_{j=1}^{5} X_{ij} = 1 \quad (i=1,2,3,4,5)$$

$$\sum_{i=1}^{5} X_{ij} = 1 \quad (j=1,2,3,4,5)$$

$$X_{ij} = 0,1 \quad (i=j=1,2,\ldots,5)$$

C. **<u>Solution Algorithm</u>**

This is an assignment problem. However, since the basic algorithm minimizes and the objective in this problem is to maximize, the matrix must be multiplied by -1:

Table 2-23

	1	2	3	4	5
1	-50	-30	-70	-40	-50
2	-60	-40	-60	-30	-60
3	-30	-60	-50	-60	-30
4	-60	-80	-40	-70	-80
5	-20	-50	-30	-80	-60

Table 2-24

	1	2	3	4	5	
1	20	40	0	30	20	-70
2	0	20	0	20	0	-60
3	30	0	10	0	30	-60
4	20	0	40	10	0	-80
5	60	30	50	0	20	-80
						-350

An optimal solution is possible:

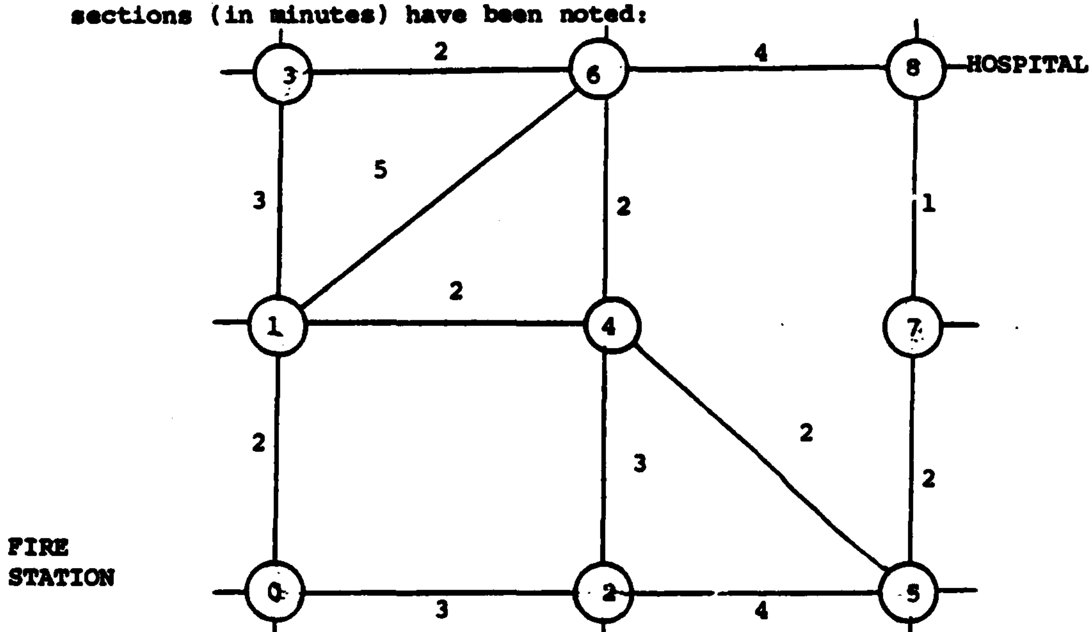
Table 2-25

	1	2	3	4	5
1	20	40	0	30	20
2	0	20	0	30	0
3	30	0	10	0	30
4	20	0	40	10	0
5	60	30	50	0	20

(1,3)	70
(2,1)	60
(3,2)	60
(4,5)	80
(5,4)	80
	350

EXAMPLE 2-7

You are trying to determine the shortest route in terms of travel time from the main fire station on base to the base hospital. You have the following partial map of the base on which the travel times between intersections (in minutes) have been noted:



Determine the quickest route from the fire station to the hospital.

Solution Procedure

This problem can be modeled and solved using the Minimal Path/Shortest Route algorithm:

Table 2-26

0	2	3	5	4	6	6	8	9
0	1	2	3	4	5	6	7	8
(0,1)2 (0,2)3	~~(1,0)2~~ (1,4)2 (1,3)3 ~~(1,6)5~~	~~(2,0)3~~ ~~(2,4)3~~ ~~(2,5)4~~	~~(3,6)2~~ ~~(3,1)3~~	~~(4,1)2~~ (4,5)2 (4,6)2 ~~(4,2)3~~	~~(5,4)2~~ (5,7)2 ~~(5,2)4~~	~~(6,3)2~~ ~~(6,4)2~~ ~~(6,8)4~~ ~~(6,1)5~~	(7,8)1 ~~(7,5)2~~	~~(8,7)1~~ ~~(8,6)4~~

1. Label 0: 0
   Cross out (1,0),(2,0)

2. (0,1): 0 + 2 = 2 ✓
   Label 1: 2
   Cross out: (3,1),(4,1),(6,1)

3. (0,2): 0 + 3 = 3 ✓
   (1,4): 2 + 2 = 4
   Label 2: 3
   Cross out: (4,2),(5,2)

4. (1,4): 2 + 2 = 4 ✓
   (2,4): 3 + 3 = 6
   Label 4: 4
   Cross out: (2,4), (5,4), (6,4)

5. (1,3): 2 + 3 = 5 ✓
   (2,5): 3 + 4 = 7
   (4,5): 4 + 2 = 6
   Label: 3: 5
   Cross out: (6,3)

6. (1,6): 2 + 5 = 7
   (2,5): 3 + 4 = 7
   (3,6): 5 + 2 = 7
   (4,5): 4 + 2 = 6 ✓
   Label 5: 6
   Cross out: (2,5),(7,5)

7. (1,6): 2 + 5 = 7
   (3,6): 5 + 2 = 7
   (4,6): 4 + 2 = 6✓
   (5,7): 6 + 2 = 8
   Label 6: 6
   Cross out: (1,6),(3,6),(8,6)

8. (5,7): 6 + 2 = 8✓
   (6,8): 6 + 4 = 10
   Label 7: 8
   Cross out: (8,7)

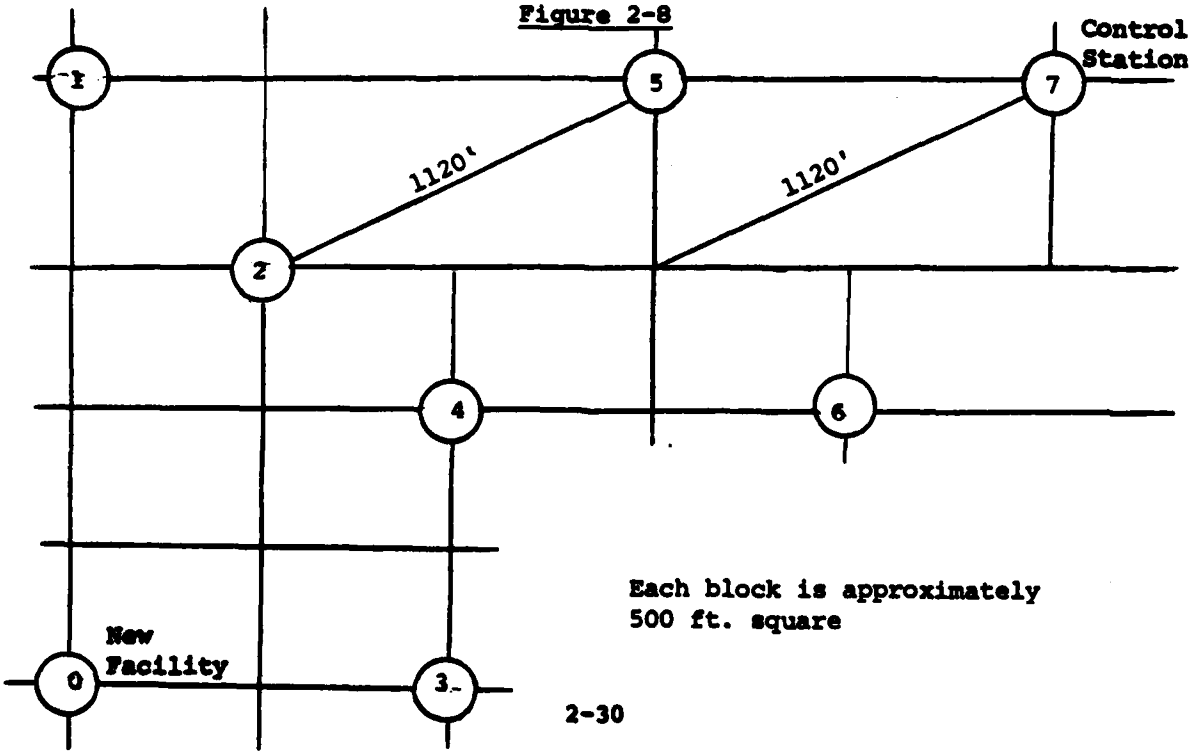
9. (6,8): 6 + 4 = 10
   (7,8): 8 + 1 = 9✓
   Label 8: 9
   Cross out: (6,8)

The optimal solution is: 8-7-5-4-1-0 or: <u>0-1-4-5-7-8 at 9 minutes</u>.

<u>EXAMPLE 2-8</u>

A facility being planned for construction at your base will be included in Energy Monitoring and Control System (EMCS). As part of the system, telephone lines must be run from the existing central monitoring and control station to the new facility. The following network represents the existing system of conduits and junction boxes through which lines can be pulled. What is the shortest route connecting the new facility to the central station?

<u>Figure 2-8</u>



Solution Procedure

This is a basic minimal path/shortest route problem in which the shortest distances between the various junction boxes are given in this table:

Table 2-27

	0	1	2	3	4	5	6	7
0	-	2000	2000	1000	2000			
1		-	1000	3000	2000	1500		
2			-	2000	1000	1120	2000	2120
3				-	1000			
4					-	1500	1000	2120
5						-	1500	1000
6							-	1500
7								-

Table 2-28

0	2000	2000	1000	2000	3120	3000	4120
0	1	2	3	4	5	6	7
(0,3) 1000 (0,1) 2000 (0,2) 2000 ~~(0,4) 2000~~	~~(1,2) 1000~~ ~~(1,5) 1500~~ ~~(1,4) 2000~~ ~~(1,3) 3000~~	~~(2,4) 1000~~ (2,5) 1120 ~~(2,3) 2000~~ ~~(2,6) 2000~~ (2,7) 2120	(3,4) 1000	(4,6) 1000 ~~(4,5) 1500~~ (4,7) 2120	(5,7) 1000 ~~(5,6) 1500~~	~~(6,7) 1500~~	

1. (0,3) = 0 + 1000 = 1000 ✓
   Label 3 @ 1000
   Cross out: (1,3),(2,3)

2. (0,1) = 0 + 2000 = 2000 ✓
   (3,4) = 1000 + 1000 = 2000 ✓
   Label both 1 and 4 @ 2000
   Cross out: (0,4),(1,4),(2,4)

3. (0,2) = 0 + 2000 = 2000 ✓
   (1,2) = 2000 + 1000 = 3000
   (4,6) = 2000 + 1000 = 3000
   Label 2 @ 2000
   Cross out: (1,2)

4. (1,5) = 2000 + 1500 = 3500
   (2,5) = 2000 + 1120 = 3120
   (4,6) = 2000 + 1000 = 3000 ✓
   Label 6 @ 3000
   Cross out: (2,6),(5,6)

5. (1,5) = 2000 + 1500 = 3500
   (2,5) = 2000 + 1120 = 3120 ✓
   (4,5) = 2000 + 1500 = 3500
   (6,7) = 3000 + 1500 = 4500
   Label 5 @ 3120
   Cross out: (1,5),(4,5)

6. (2,7) = 2000 + 2120 = 4120 ✓
   (4,7) = 2000 + 2120 = 4120 ✓
   (5,7) = 3120 + 1000 = 4120 ✓
   (6,7) = 3000 + 1500 = 4500
   Label 7 @ 4120
   Cross out: (6,7)

Alternative shortest routes:

0 - 2 - 5 - 7

0 - 3 - 4 - 7

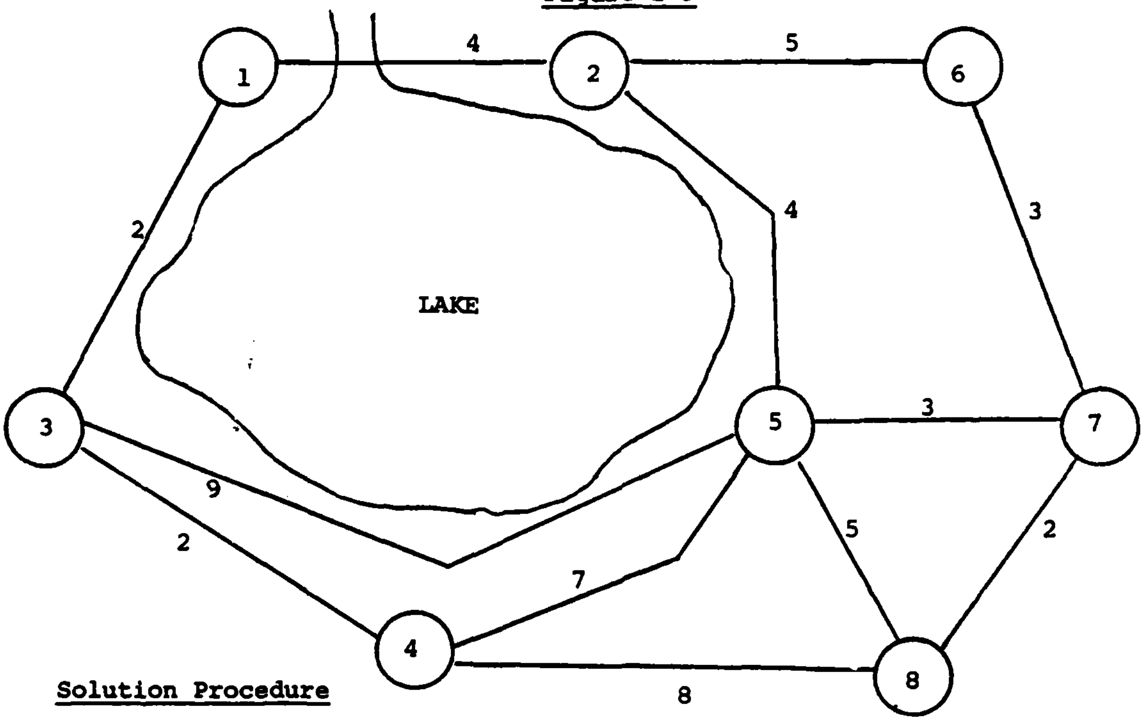
0 - 2 - 7

Z = 4120 ft.

EXAMPLE 2-9

The Base Civil Engineer is planning the development of a recreation area surrounding a lake located on the base. The planners/site developers have identified ideal locations in the park for a lodge, picnic grounds, a boat dock, camping areas, recreation fields, and other facilities. These locations are represented by nodes on the network illustrated below. The branches of the network represent possible road alternatives in the recreation area. If the park designers want to minimize the total road miles that must be constructed in the park and still permit access to all facilities, which road alternatives should be constructed:

Figure 2-9



Solution Procedure

This is a minimal spanning tree problem:

Table 2-29

√	√	√	√	√	√	√	√
1	2	3	4	5	6	7	8
(1,3)2 (1,2)4	~~(2,1)4~~ (2,5)4 (2,6)5	~~(3,1)2~~ (3,4)2 (3,5)9	~~(4,3)2~~ (4,5)7 (4,8)8	(5,7)3 ~~(5,2)4~~ (5,8)5 (5,4)7 (5,3)9	~~(6,7)3~~ (6,2)5	(7,8)2 ~~(7,5)3~~ (7,6)3	~~(8,7)2~~ (8,5)5 (8,4)8

1. Arbitrarily start at node (1)
   Label (✓) node (1)
   Connect 1-3
   Label node (3)
   Cross out (3,1)

2. (1,2) 4
   (3,4) 2✓
   Connect 3-4
   Cross out (4,3)

3. (1,2) 4✓
   (3,5) 9
   (4,5) 7
   Connect 1-2
   Cross out (2,1)

4. (2,5) 4✓
   (3,5) 9
   (4,5) 7
   Connect 2-5
   Cross out (5,2)

5. (2,6) 5
   (3,5) 9
   (4,5) 7
   (5,7) 3✓
   Connect 5-7
   Cross out (7,5)

6. (2,6) 5
   (3,5) 9
   (4,5) 7
   (5,8) 5
   (7,8) 2✓
   Connect 7-8
   Cross out (8,7)

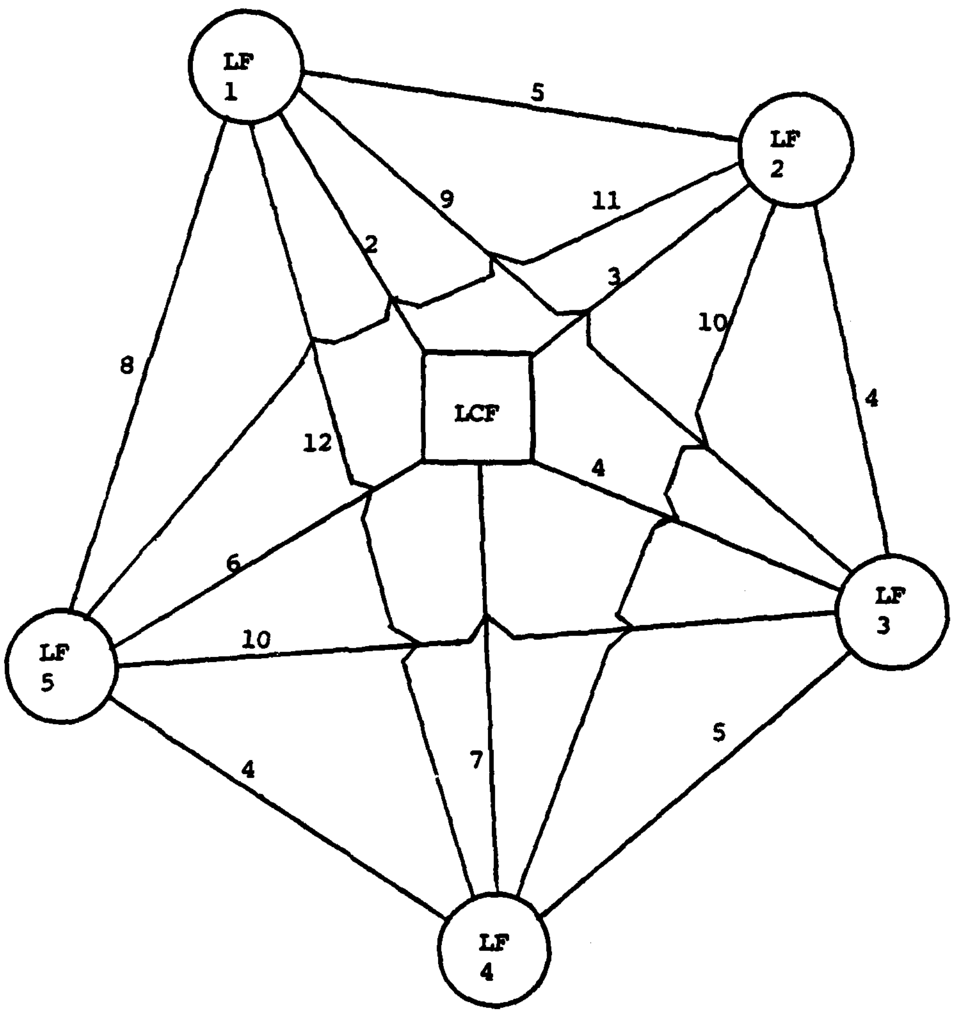
7. (2,6) 5
   (3,5) 9
   (4,5) 7
   (5,8) 5
   (7,6) 3✓
   (8,5) 5
   Connect 7-6
   Cross out (6,7)

Total spanning tree length:		
	1-3	2
	3-4	2
	1-2	4
	2-5	4
	5-7	3
	7-8	2
	7-6	3
		20

EXAMPLE 2-10

One configuration of the MX missile deployment system being studied is illustrated by this network diagram:

Figure 2-10



Each Launch Control Facility (LCF) directs the movement of a mobile launcher through a tunnel system linking 5 Launch Facilities (LF). Find the minimal spanning tree representing tunnel links connecting the LCF and LFs.

Solution Procedure

Table 2-30

√	√	√	√	√	√
LCF(0)	LF-1	LF-2	LF-3	LF-4	LF-5
(0,1) 2	~~(1,0) 2~~	~~(2,0) 3~~	(3,0) 4	(4,5) 4	~~(5,0) 6~~
(0,2) 3	~~(1,2) 5~~	~~(2,3) 4~~	~~(3,2) 4~~	~~(4,3) 5~~	~~(5,4) 4~~
(0,3) 4	~~(1,5) 8~~	~~(2,1) 5~~	(3,4) 5	~~(4,0) 7~~	~~(5,1) 8~~
~~(0,5) 6~~	~~(1,3) 9~~	~~(2,4) 10~~	~~(3,1) 9~~	~~(4,2) 10~~	~~(5,3) 10~~
~~(0,4) 7~~	~~(1,4) 12~~	~~(2,5) 11~~	~~(3,5) 10~~	~~(4,1) 12~~	~~(5,2) 11~~

1. Label (√) LCF(0)

2. Connect (0-1)
   Label 1
   Cross out: (1,0)

3. (0,2): 3√
   (1,2): 5
   Connect (0-2)
   Label 2
   Cross out: (1,2), (2,0), (2,1)

4. (0,3): 4√
   (1,5): 8
   (2,3): 4
   Connect 3 from either 0 or 2 (arbitrarily select 0)
   Label 3
   Cross out: (1,3), (2,3), (3,2), (3,1)

5. (0,5) 6
   (1,5) 8
   (2,4) 10
   (3,4) 5√
   Connects (3-4)
   Label 4
   Cross out: (0,4), (1,4), (2,4), (4,0), (4,3), (4,2), (4,1)

6. (0,5) 6
   (1,5) 8
   (2,5) 11
   (3,5) 10
   (4,5) 4✓
   Connect 4-5
   Label 5
   Cross out: (0,5), (1,5), (2,5), (3,5), (5,0), (5,4), (5,0), (5,4), (5,1), (5,3), (5,2)

Minimal Spanning Tree:

0 - 1	2
0 - 2	3
0 - 3	4
3 - 4	5
4 - 5	4
	18

EXAMPLE 2-11

An independent testing lab has been retained by the base civil engineer to accomplish materials testing for a major on-base construction project. Five separate tests are to be accomplished. These tests can be accomplished in any sequence. However, the time required to set up the test equipment depends on the sequence in which the tests are conducted. Equipment set-up times (from test i to test j) are as follows:

Table 2-31

	i \ j	TO TEST 1	2	3	4	5
FROM TEST	1	-	9	12	9	17
	2	3	-	6	3	14
	3	8	2	-	13	5
	4	6	14	7	-	4
	5	12	3	9	1	-

In what sequence should the tests be accomplished?

Solution Procedure

This problem can be formulated as a <u>closed-circuit model</u>.

Table 2-32

i \ j	1	2	3	4	5	
1	M	0	0	0	8	9
2	0	M	0	0	11	3
3	6	0	M	11	3	2
4	2	10	0	M	0	4
5	11	2	5	0	M	1
			3			22

Table 2-33

i \ j	1	2	3	4	5
1	M	0 / 0	0 / 0	0 / 0	8
2	2 / 0	M	0 / 0	0 / 0	11
3	6	3 / 0	M	11	3
4	2	10	0 / 0	M	3 / 0
5	11	2	5	2 / 0	M

Figure 2-11

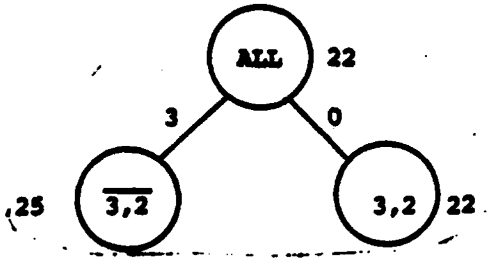


Table 2-34

i \ j	1	3	4	5
1	M	0^{0}	0^{0}	8
2	0^{2}	M	0^{0}	11
4	2	0^{0}	M	0^{8}
5	11	5	0^{5}	M

Figure 2-12

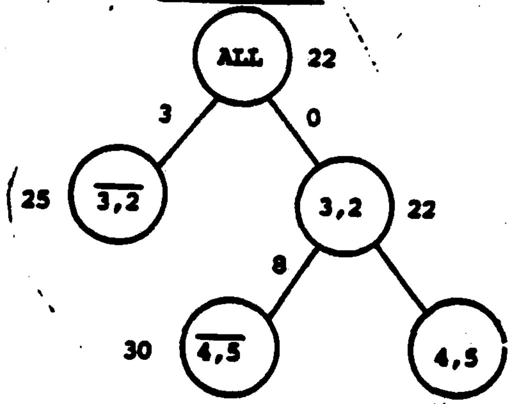


Table 2-35

i \ j	1	3	4	
1	M	0^{0}	0^{0}	
2	0^{6}	M	0^{0}	
5	6	0^{6}	M	5

Figure 2-13

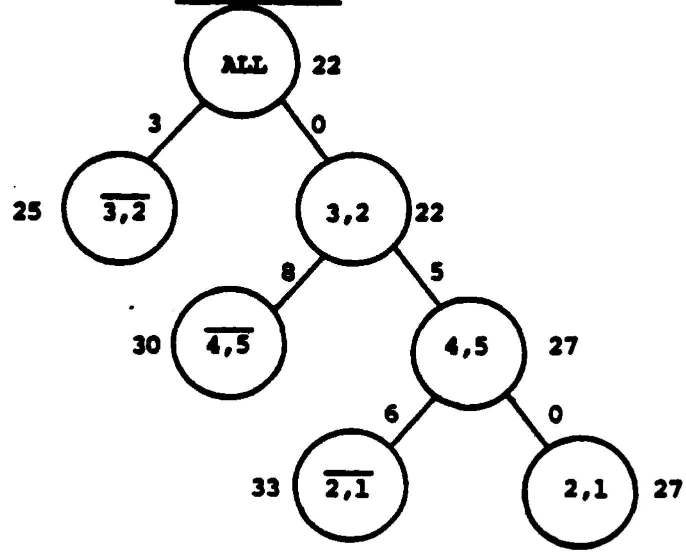


Table 2-36

i \ j	3	4
1	M	0 M
5	0 M	M

Figure 2-14

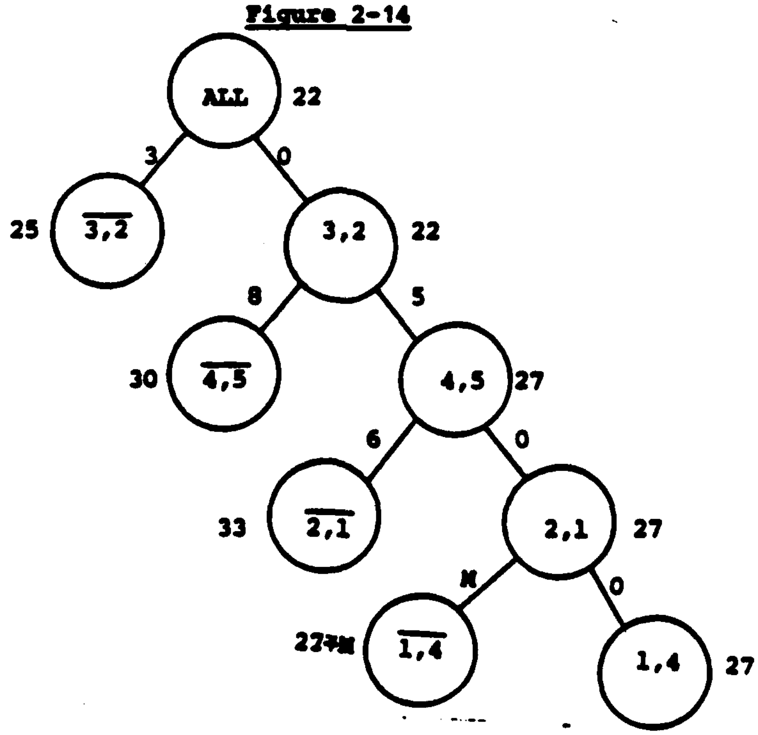


Table 2-37

i \ j	3
5	0

The left node (3,2) has a bound of 25 which is less than the bound on (5,3).

Table 2-38

i \ j	1	2	3	4	5	
1	M	2 0	0 0	0 0	8	
2	2 0	M	0 0	0 0	11	
3	3	M	M	8	3 0	3
4	2	10	0 0	M	0 0	
5	11	2	5	2 0	M	

Figure 2-15

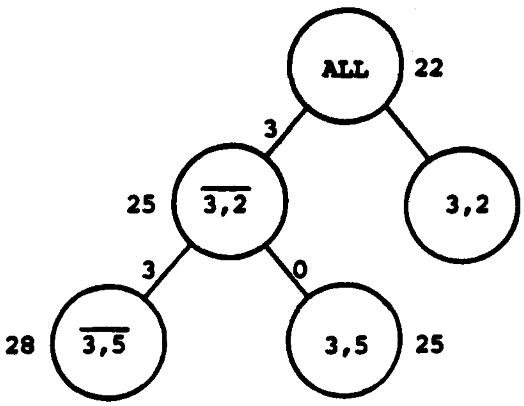


Table 2-39

i \ j	1	2	3	4
1	M	2 0	0 0	0 0
2	2 0	M	0 0	0 0
4	2	10	0 0	M
5	11	2	M	2 0

Table 2-40

i \ j	1	3	4
2	2 0	0 0	0 0
4	2	2 0	M
5	11	M	11 0

Figure 2-16

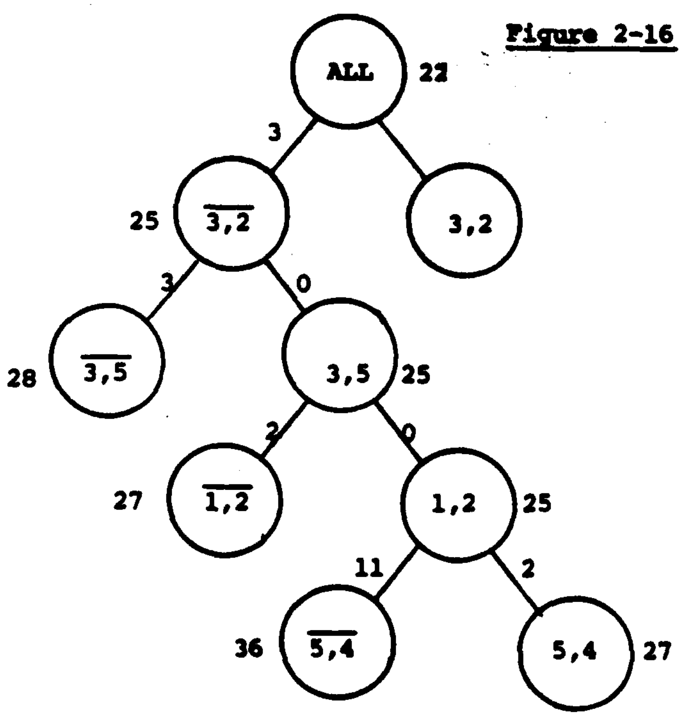


Table 2-41

i \ j	1	3	
2	0	0	
4	0	M	2

The bound on node (5,4) is equal to the bound on the solution previously determined.

The optimal solution is: 1-4-5-3-2-1 at 9+4+9+2+3=<u>27</u>

EXAMPLE 2-12

The Structural Maintenance and Repair Team (SMART) accomplishes scheduled maintenance on high-use facilities. To support the SMART, a shop trailer containing tools and bench stock is moved between six work zones. The following matrix tabulates the time it takes to relocate and set up the trailer:

Table 2-42

TO ZONE j:

FROM ZONE i: i \ j	1	2	3	4	5	6
1	-	30	45	60	45	20
2	35	-	40	30	15	20
3	40	35	-	20	30	45
4	60	25	25	-	15	30
5	50	20	35	20	-	15
6	15	25	40	30	20	-

In what sequence should the SMART crew be routed through the six work zones to minimize the total relocation and set-up time?

Solution Procedure

This problem can be formulated as a closed circuit problem:

Table 2-43

	1	2	3	4	5	6	
1	M	5	15	40	25	5 0	20
2	20	M	15	15	5 0	5	15
3	20	10	M	15 0	10	25	20
4	45	5	15 0	M	0 0	25	15
5	35	5 0	10	5	M	0	15
6	25 0	5	15	15	5	M	15
		5	10				115

Figure 2-17

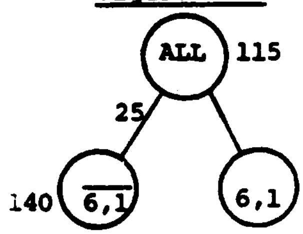


Table 2-44

	2	3	4	5	6	
1	10 0	10	35	20	M	5
2	M	15	15	5 0	5	
3	10	M	15 0	10	25	
4	5	10 0	M	0 0	15	
5	0 0	10	5	M	5 0	

Figure 2-18

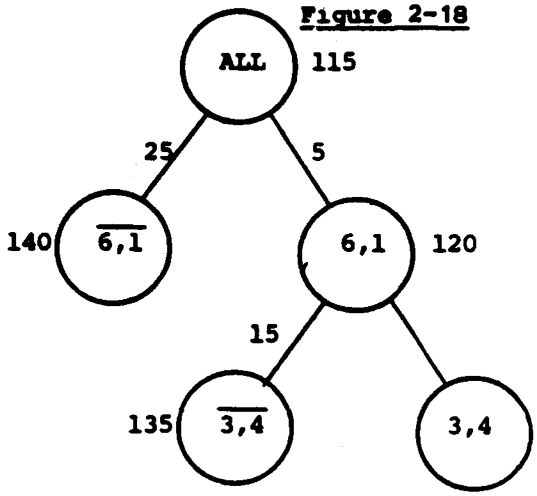


Table 2-45

	2	3	5	6
1	0 0	0 0	20	M
2	M	5	5 0	5
4	5	M	5 0	15
5	0 0	0 0	M	5 0

Figure 2-19

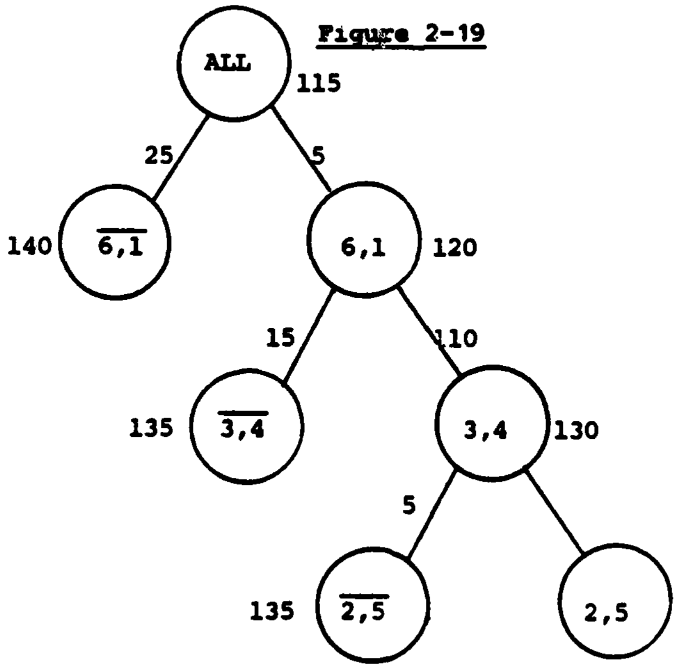


Table 2-46

	2	3	6	
1	0 0	0 0	M	
4	10 0	M	10	5
5	M	0 0	10 0	

Figure 2-20

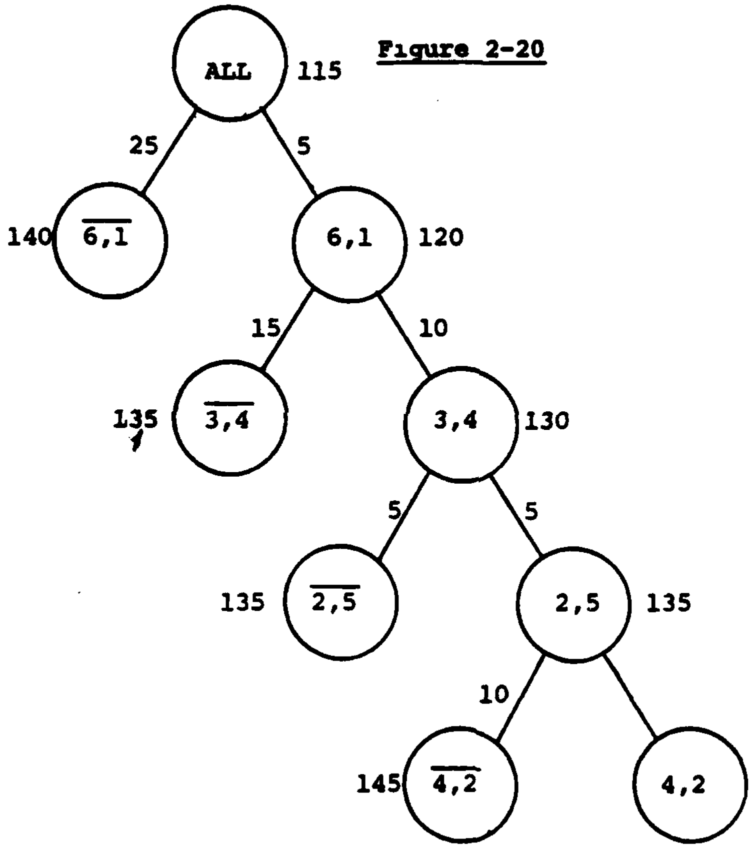
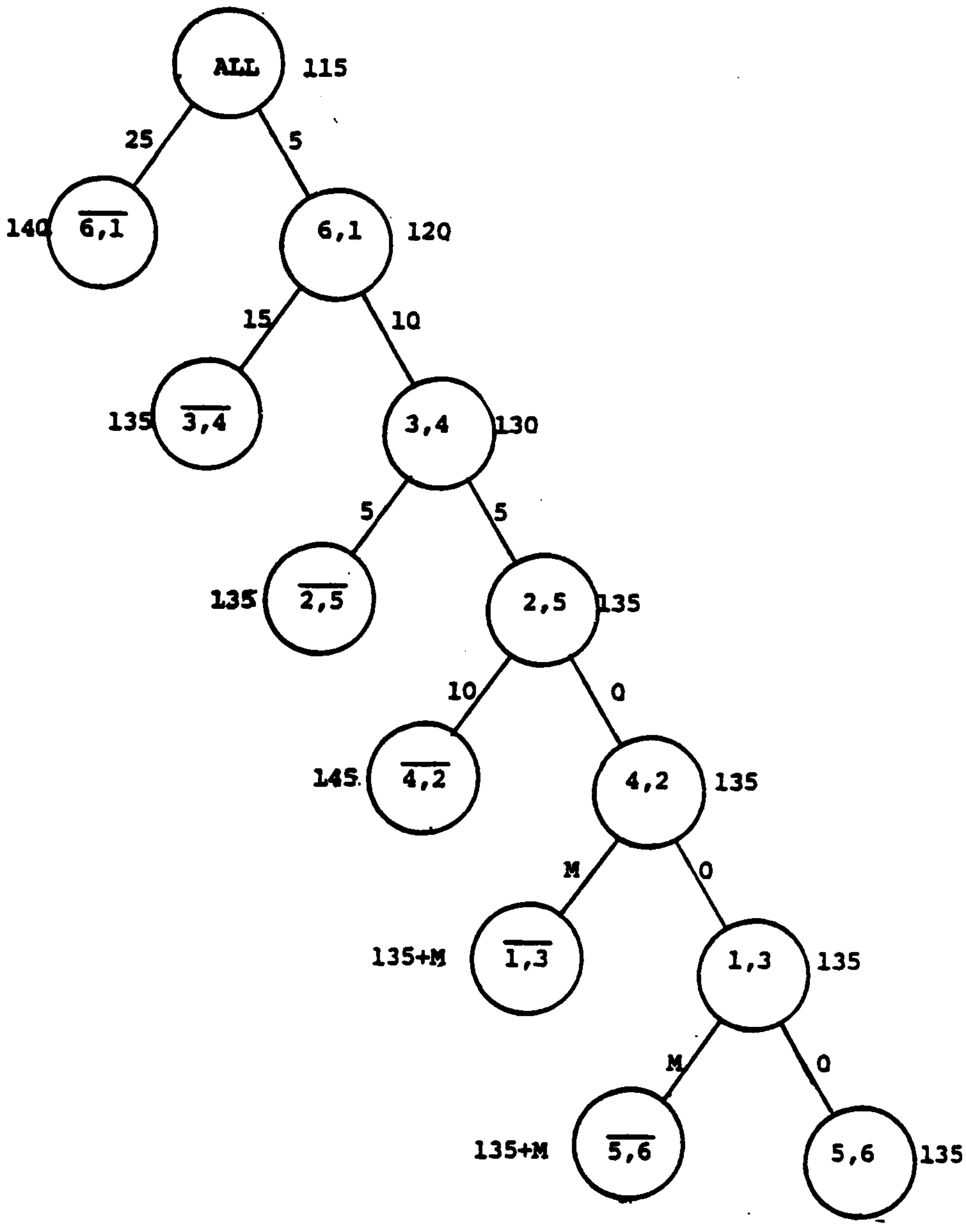


Table 2-47

	3	6
1	M 0	M
5	M	M 0

Figure 2-21



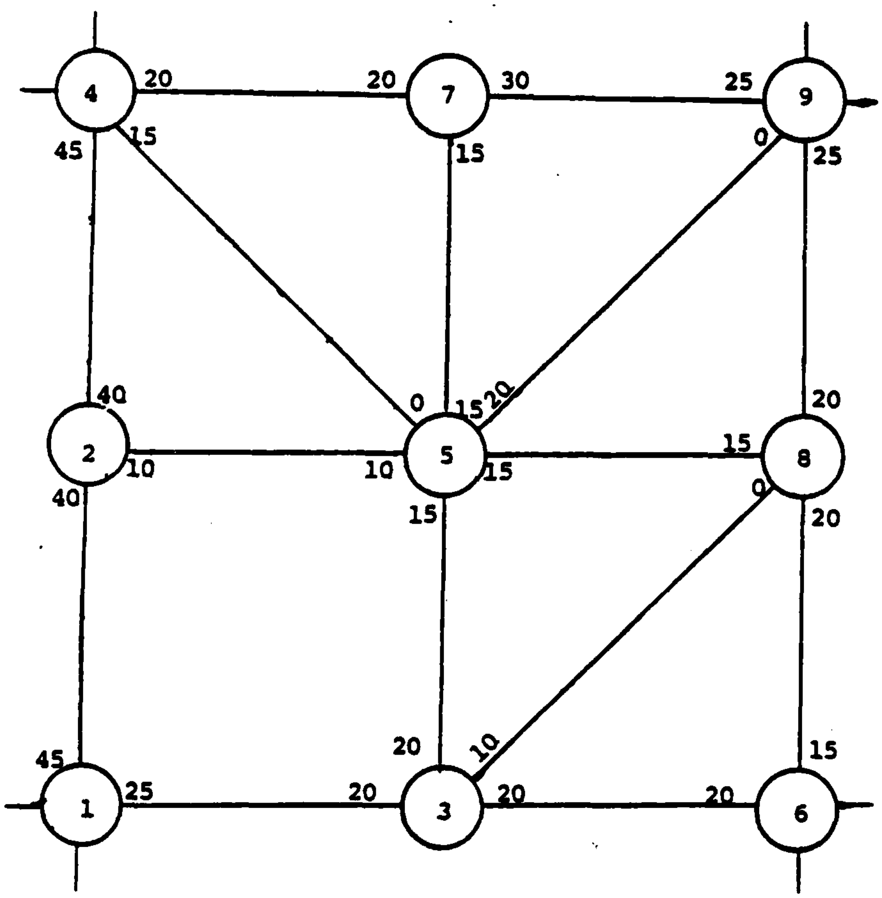
OPTIMAL SEQUENCE: 1-3-4-2-5-6-1

Z = 135

EXAMPLE 2-13

The following capacitated network represents a portion of the base road system. The capacities on the network represent the maximum average vehicle traffic/minute as determined in a recent traffic survey.

Table 2-48

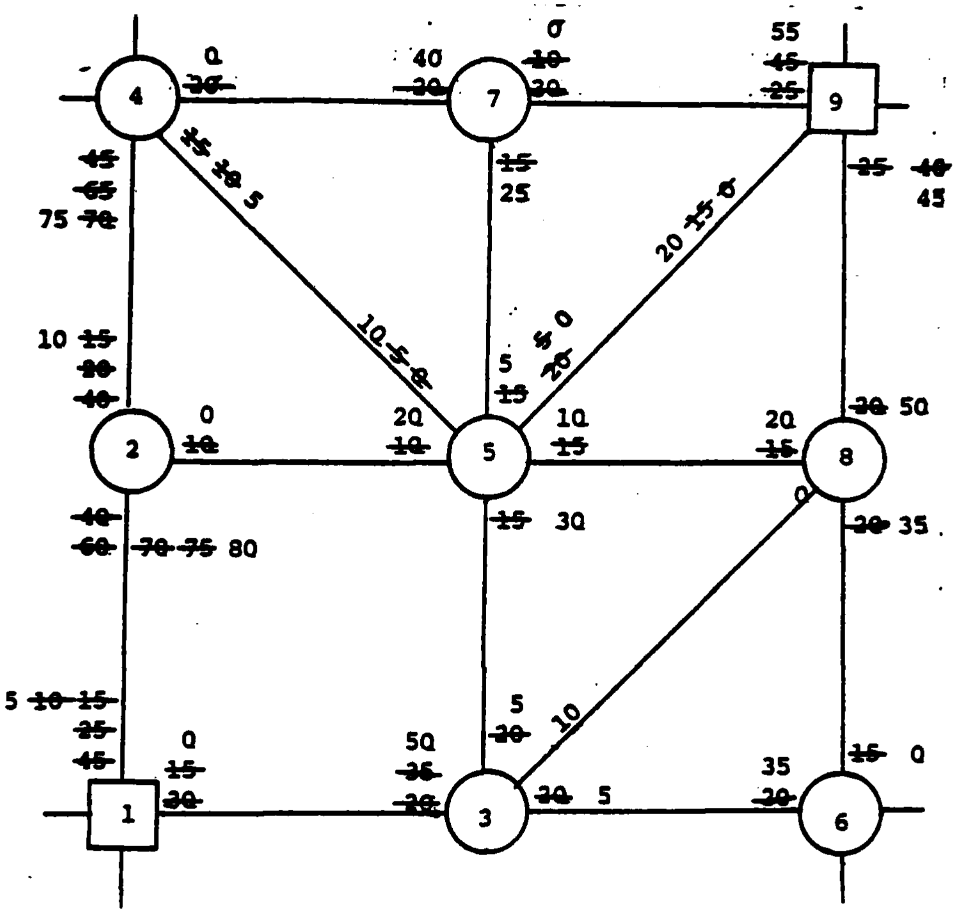


What is the total volume of traffic that can be carried through this system, assuming traffic enters the network at intersection (1) and departs through intersection (9)?

Solution Procedure

This can be modeled as a <u>maximal flow</u> problem.

Table 2-49



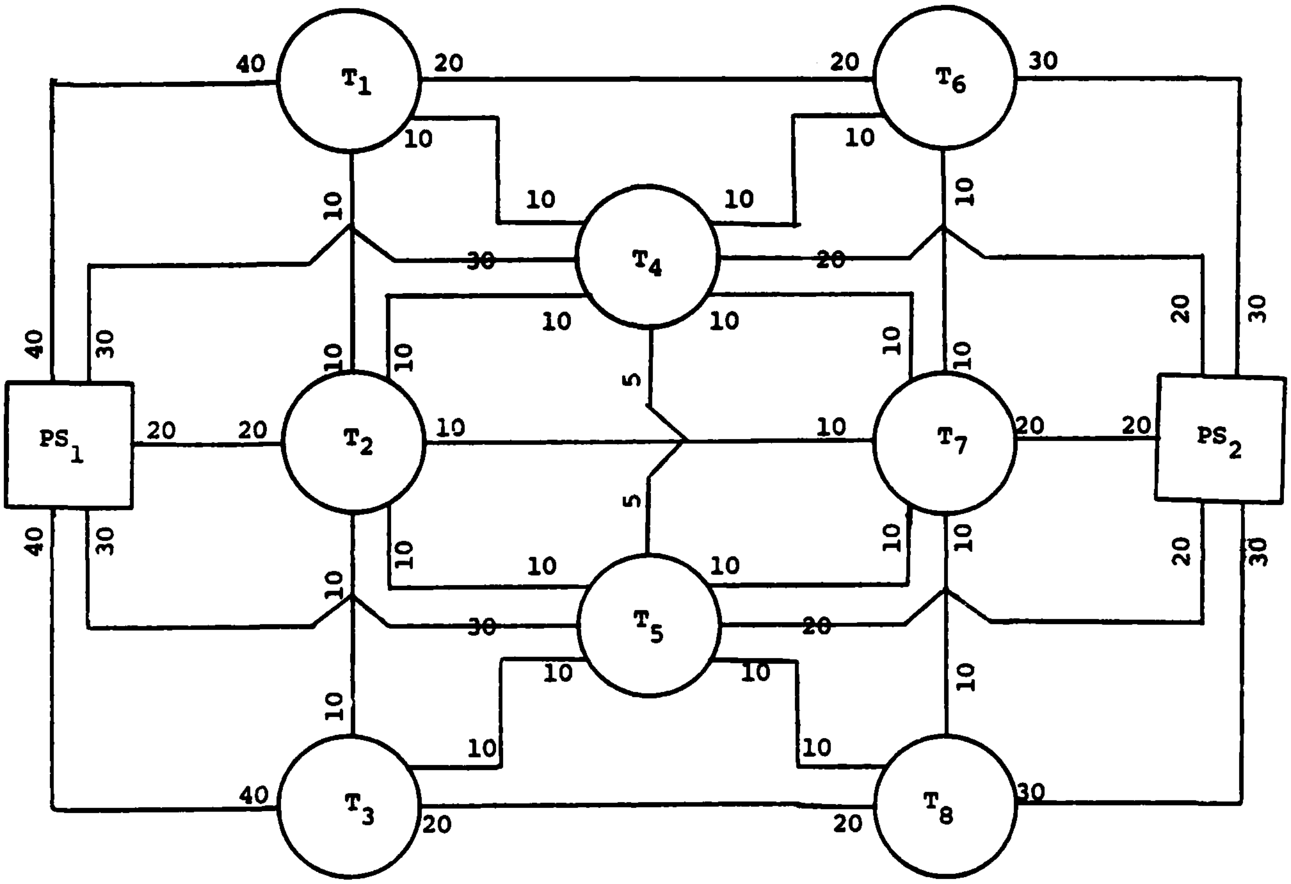
20
15
15
10
5
5

1 - 2 - 4 - 7 - 9	20
1 - 3 - 6 - 8 - 9	15
1 - 3 - 5 - 9	15
1 - 2 - 5 - 7 - 9	10
1 - 2 - 4 - 5 - 9	5
1 - 2 - 4 - 5 - 8 - 9	5
	70

EXAMPLE 2-14

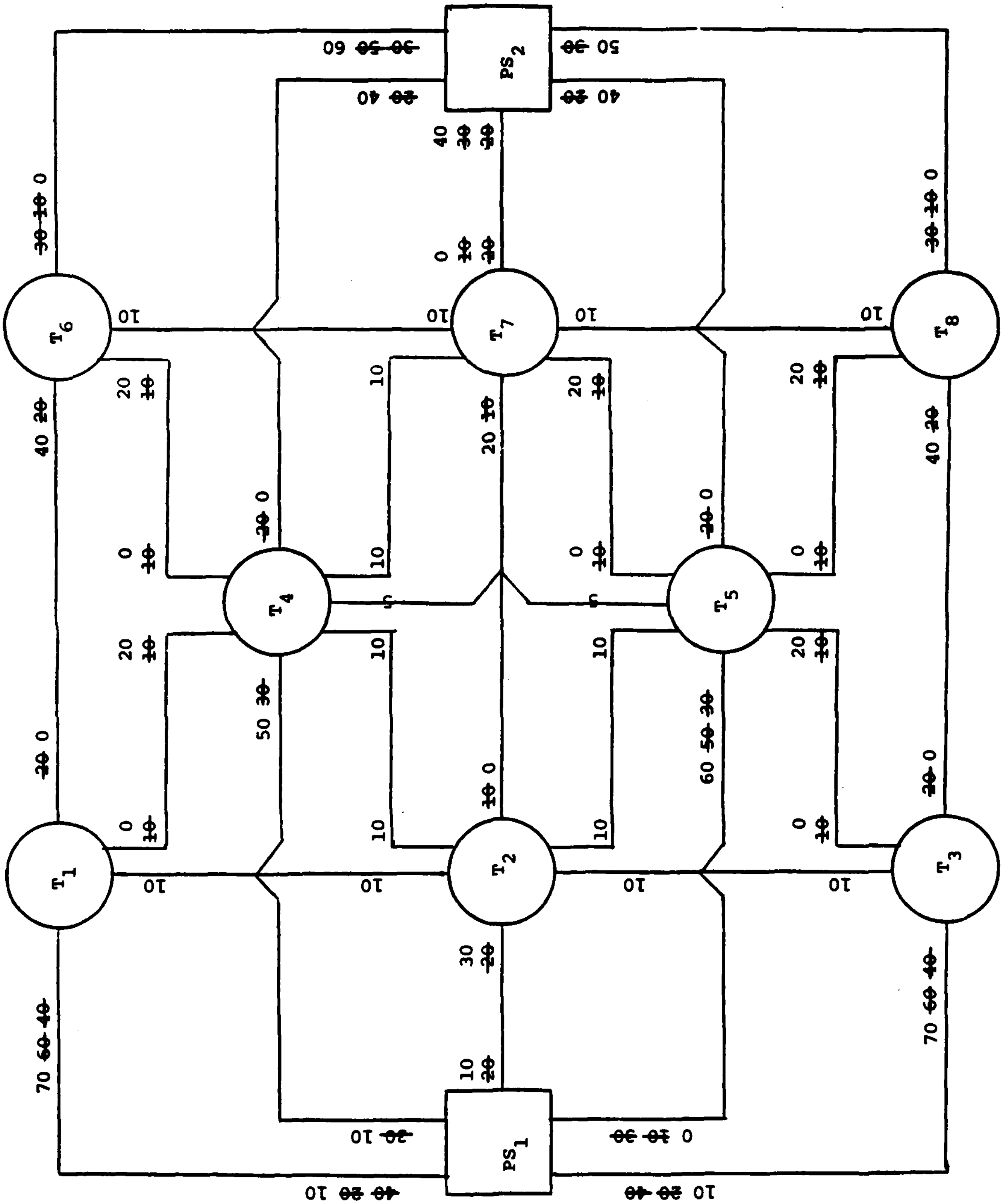
The following diagram illustrates the proposed layout of a POL tank farm to be constructed at a particular base:

Figure 2-22



Determine the maximum flow of product that can be carried through the system from Pump Station 1 to Pump Station 2.

Figure 2-23



$PS_1 - T_1 - T_6 - PS_2$	20
$PS_1 - T_3 - T_8 - PS_2$	20
$PS_1 - T_4 - PS_2$	20
$PS_1 - T_5 - PS_2$	20
$PS_1 - T_2 - T_7 - PS_2$	10
$PS_1 - T_1 - T_4 - T_6 - PS_2$	10
$PS_1 - T_3 - T_5 - T_8 - PS_2$	10
$PS_1 - T_5 - T_7 - PS_2$	10
	120

EXAMPLE 2-15

The Base Civil Engineer has given you the opportunity to develop and manage the accomplishment of a small project requested by the base commander. Although not complex, the project is time-critical and must be closely controlled. After analyzing the project system and discussing it with various individuals who will be involved in the project, you have compiled the following information:

Table 2-50

ACTIVITY (i - j)	a	m	b
0 - 1	3	7	11
1 - 3	2	2.5	6
1 - 5	2	3	4
0 - 2	6	7	14
2 - 3	2	3	4
2 - 4	2.5	3	3.5
4 - 5	2.5	4	5.5
0 - 3	4.5	5.5	9.5
3 - 4	1	2	3
3 - 5	1	2	3

a = optimistic time m = most likely time b = pessimistic time

1. Construct the project system network. Assume event 0 is the origin and event 5 is the terminal node.

2. Compute the expected time (t_e) and variance for each activity.

3. Find the critical path.

4. Construct a graph of the probability of completion (Y - axis) versus the completion time in days (X - axis). The domain should vary between 14 days and 20 days.

5. The Base Civil Engineer wants to give the base commander an estimated completion date in which the confidence level is at least 98.0%. How many days should be scheduled for completing this project at the desired confidence level?

Solution Procedure

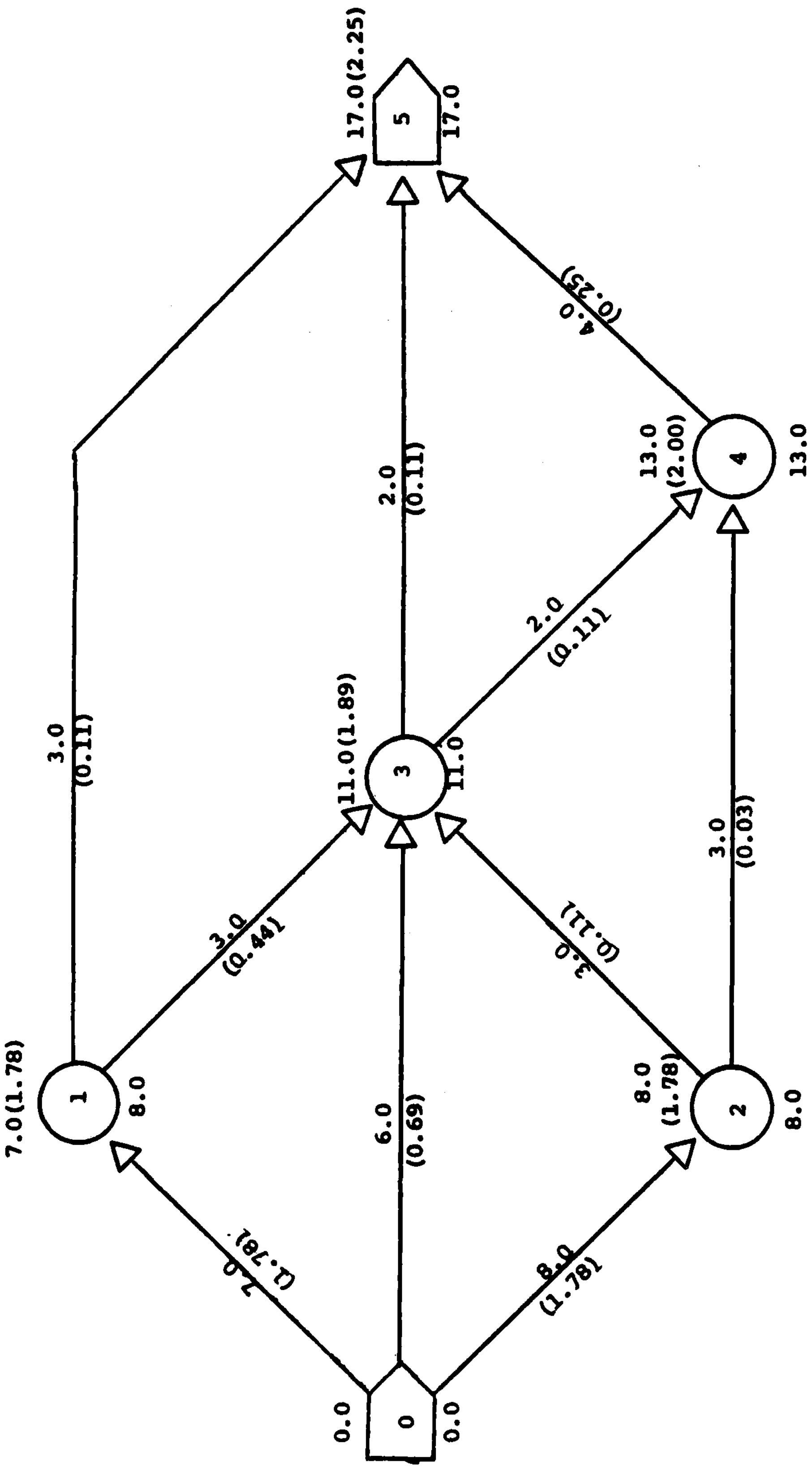
$$t_e = \frac{(a + 4m + b)}{6}$$

$$v = \frac{(b-a)^2}{36}$$

Table 2-51

ACT.	a	m	b	t_e	v
0 - 1	3	7	11	7.0	1.78
1 - 3	2	2.5	6	3.0	0.44
1 - 5	2	3	4	3.0	0.11
0 - 2	6	7	14	8.0	1.78
2 - 3	2	3	4	3.0	0.11
2 - 4	2.5	3	3.5	3.0	0.03
4 - 5	2.5	4	5.5	4.0	0.25
0 - 3	4.5	5.5	9.5	6.0	0.69
3 - 4	1	2	3	2.0	0.11
3 - 5	1	2	3	2.0	0.11

Figure 2-24



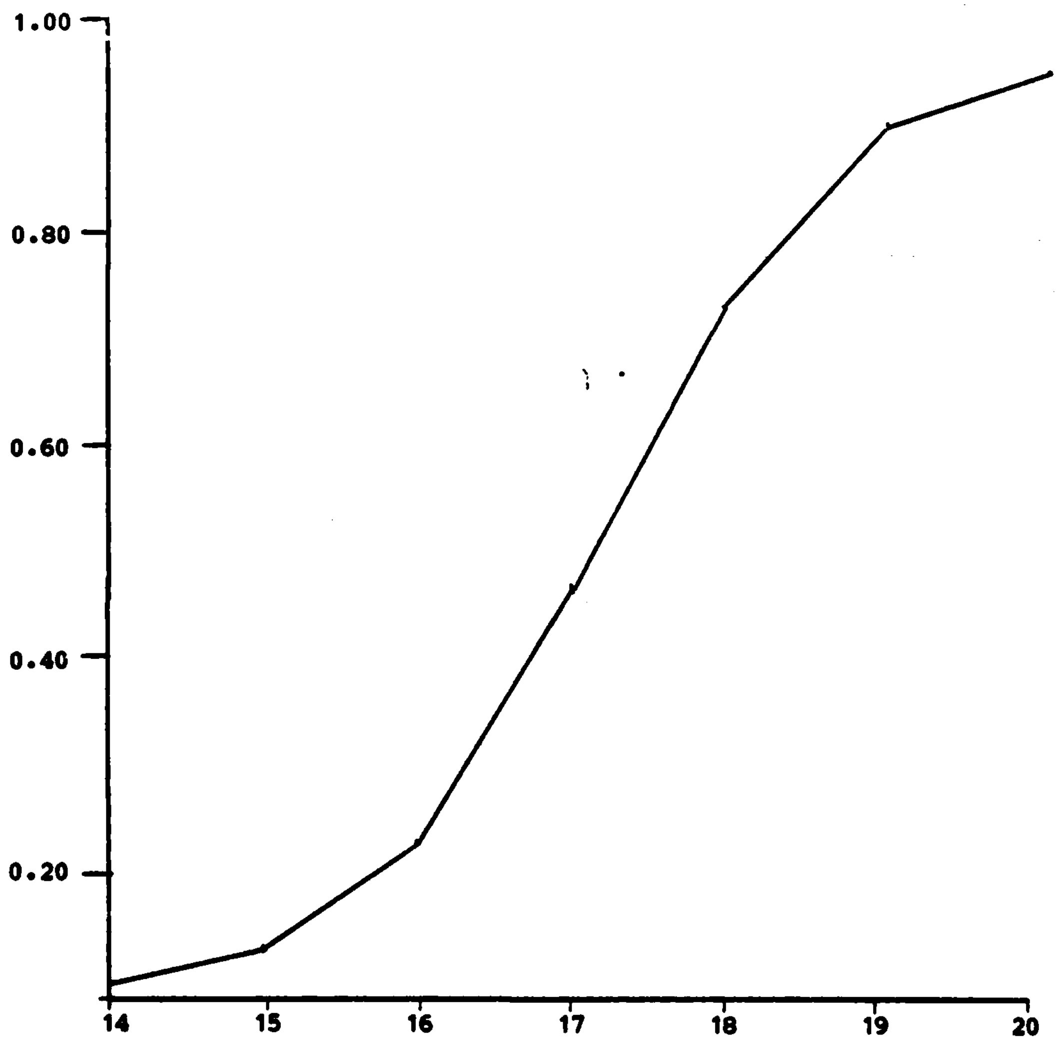
CRITICAL PATH: 0 - 2 - 3 - 4 - 5 @ 17.0(2.25)

$$P(T_5 \leq E(T_5)) = P\left(Z_5 \leq \frac{T_5 - E(T_5)}{\sqrt{v}}\right)$$

$$\sigma = \sqrt{v} = \sqrt{2.25} = 1.50$$

T_5	Z	TABLE	$P(T_5<E(T_5)$
14	-2.00	0.4772	0.0228
15	-1.33	0.4082	0.0918
16	-0.67	0.2486	0.2514
17	0.00	0.0	0.5000
18	0.67	0.2486	0.7486
19	1.33	0.4082	0.9082
20	2.00	0.4772	0.9772

Figure 2-25



$P(T_5 \leq 20) = 0.9772$

$P(T_5 \leq 21)$: $\frac{(21-17)}{1.5} = \frac{4.0}{1.5} = 2.67$ (.4962)

$P(T_5 \leq 21) = 0.9962$

You need to schedule at least 21 days for a confidence level $\geq$ 98.0%.

Alternatively:

$$\frac{(X-17)}{1.5} = 2.055 \qquad (2.055 = Z \leq 98.0\%)$$

$$X = 20.08 \text{ days}$$

EXAMPLE 2-16

As the engineer on a small construction project, you have compiled the following data:

Table 2-52

EVENT	ACT.	a	m	b	a*	COST
0						
1	0 - 1	5	7	9	3	120
2	0 - 2	4	4	5	3	140
3	0 - 3	6	9	11	5	50
	1 - 3	3	3	6	2	170
	2 - 3	3	3	3	-	-
4	1 - 4	5	6	8	4	180
	3 - 4	1	1	1	-	-
5	2 - 5	6	7	7	4	60
	3 - 5	2	2	2	-	-
6	3 - 6	4	5	5	3	30
	4 - 6	3	4	7	2	60
	5 - 6	2	3	4	1	20

a* represents the optimistic completion time estimate when the task is "crashed" or expedited at the associated cost. Determine the greatest probability completing the project in 18.0 days (or less) if $500 is available for expediting the project.

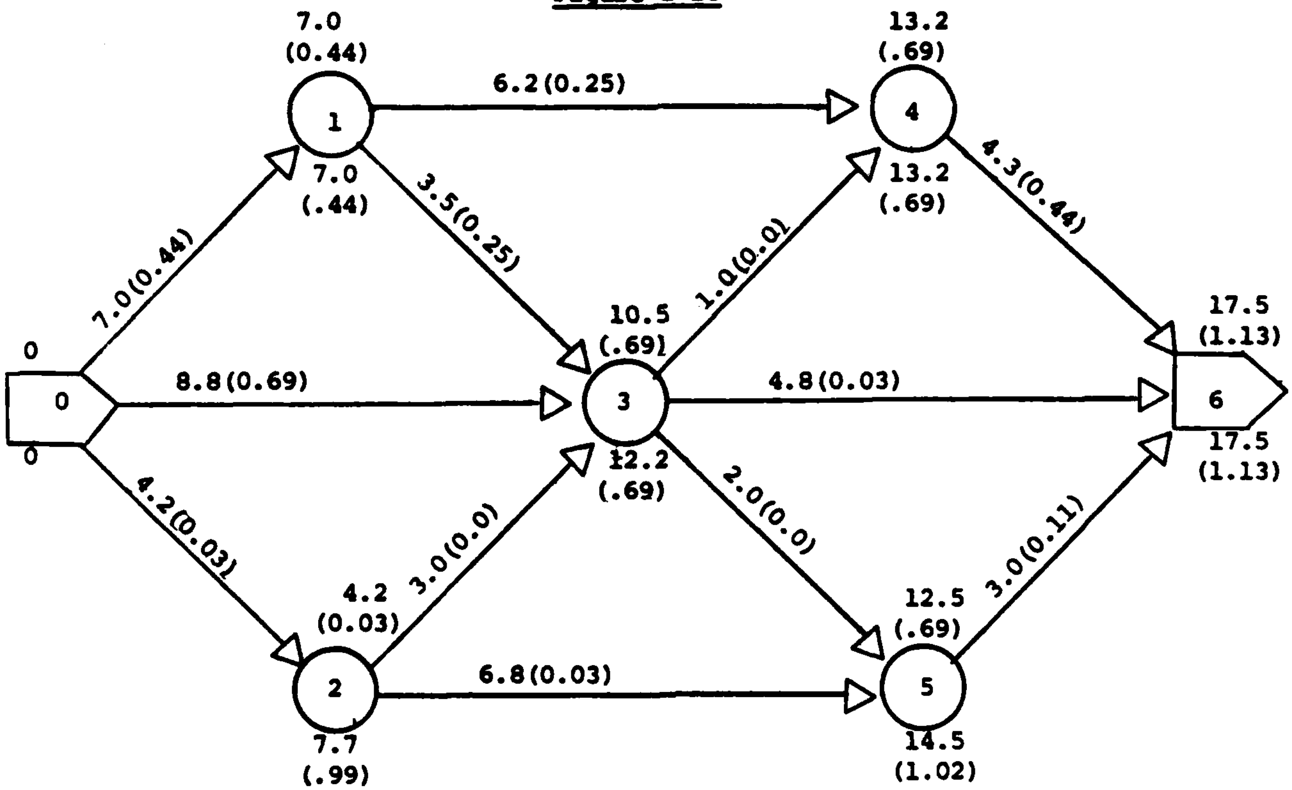
Solution Procedure

$t_e = \frac{(a+4m+b)}{6}$ $v = \frac{(b-a)^2}{36}$

Table 2-54

ACT.	t_e	v	t_e^*	v^*	$\frac{\text{COST}}{t_e - t_e^*}$
0 - 1	7.0	0.44	6.7	1.00	400
0 - 2	4.2	0.03	4.0	0.11	700
0 - 3	8.8	0.69	8.7	1.00	500
1 - 3	3.5	0.25	3.3	0.44	800
2 - 3	3.0	0.0	-	-	-
1 - 4	6.2	0.25	6.0	0.44	900
3 - 4	1.0	0.0	-	-	-
2 - 5	6.8	0.03	6.5	0.25	200
3 - 5	2.0	0.0	-	-	-
3 - 6	4.8	0.03	4.7	0.11	300
4 - 6	4.3	0.44	4.2	0.69	600
5 - 6	3.0	0.11	2.8	0.25	100

Figure 2-26



CRITICAL PATH: 0 - 1 - 4 - 6 @ 17.5 (1.13)

This path remains critical when all activities are crashed. Consequently, it is only necessary to crash activities 0-1, 1-4, and 4-6, at a cost of 120 + 180 + 60 = 360 to achieve the fully-crashed t_e* of <u>16.9</u> and associated variance of <u>2.13</u>. If an amount less than 360 was available for crashing, the activities should be expedited in order of increasing crash ratio:

Activity	Cost / $t_e - t_e$*
0 - 1	400
4 - 6	600
1 - 4	900

$$Pr(t_e \leq 18) = \frac{18-16.9}{\sqrt{2.13}} = \frac{1.10}{1.46} = \frac{1.10}{1.46} = +0.75$$

$$Pr = 0.5000 + 0.2734 = 0.7734$$

If the project is not crashed, this probability is:

$$\frac{18.0-17.5}{\sqrt{1.13}} = \frac{0.5}{1.065} = 0.47$$

$$Pr = 0.5000 + 0.1808 = 0.6808$$

Consequently, the probability of completing the project in 18.0 days or less can be increased from 0.68 to 0.77 by spending $360 to crash critical activities 0-1, 1-4, and 4-6.

<u>EXAMPLE 2-17</u>

A Base Civil Engineer has been tasked to make an economic analysis to determine if the siding on a building should continue to be painted or if vinyl siding should be applied. The following data is available:

a. <u>Painting</u>: O&M costs for the first year are $500; these costs increase at $60 per year for 20 years. Major repair during 5th year of $2,000, during the 10th year of $3,000, and during the 15th year of $4,000.

b. <u>Vinyl Siding</u>: Initial cost of $15,000. No O&M Costs.

Which alternative is more economical, assuming a rate of return of 10%?

<u>Solution Procedure</u>

a. Using the <u>equivalent annual cost method</u>: let AC = equivalent annual cost per year, A/G = the factor to convert a gradient series to an equivalent uniform annual series for n periods (looked-up in appropriate table), P/F = present worth factor at rate i for period n (from table), A/P = uniform series capital recovery factor at rate i for n periods (from table)

(1) Painting:

AC_1	= $500.00
AC_2 = \$60 (A/G, i=10%, n=20) = \$60 (6.51)	= 390.00
AC_3 = \$2000 (P/F, i=10%, n=5) · (A/P, i=10%, n=20) = \$2000 (0.6209) · (0.11746)	= 145.86
AC_4 = \$3000 (P/F, 10%, n=10) · (A/P, i=10%, n=20) = \$3000 (0.3855) · (0.11746)	= 135.84
AC_5 = \$4000 (P/F, 10%, n=15) · (A/P, i=10%, n=20) = \$4000 (0.2394) · (0.11746)	= 112.48
TOTAL	= $1284.78

(2) Vinyl Siding:

AC = \$15,000 (A/P, i=10%, n=20)
= \$15,000 (0.11746) = $1761.90

b. Using the present value method:

(1) Painting:

\$1284.78 (P/A, i=10%, n=20)
= 1284.78 (8.514) = $10,938.62

(2)

\$1761.90 (P/A, i=10%, n=20)
= 1761.90 (8.514) = $15,000.82

EXAMPLE 2-18

A Base Civil Engineer, prompted by the Environmental Protection Agency and others, must construct a new sewage disposal facility. The base is projected to grow, and he knows future expansion of the facility is inevitable. The question arises whether to design and build the facility to accommodate added capacity at a later date. To do this would add $20,000 to the present cost.

Future expansion will cost $100,000. However, if the accommodation is not made during the current construction, future expansion will cost $140,000. Assume the life of the treatment plant to be 50 years. Maintenance costs will not be affected with or without the accommodation.

How soon must the expansion be required in order to justify the additional $20,000 expenditure now? Assume i=10%.

Solution Procedure

\$20,000 spent now would save \$40,000 n years from now. Therefore:

$$(P/F,\ i=10\%,\ n) = \frac{P}{F} = \frac{20{,}000}{40{,}000} = 0.5000$$

From tables: $n \approx 7.2$ years.

Alternatively:

$$20{,}000 = 40{,}000 \cdot \left[\frac{1}{(1.10)^n}\right]$$

$$(1.10)^n = \frac{40{,}000}{20{,}000} = 2.0$$

$$\log(1.10) = 0.09531$$

$$\log(2.0) = 0.69315$$

$$\therefore n\,(0.09531) = 0.69315$$

$$n = 0.69315/0.09531 = 7.27 \text{ years.}$$

CHAPTER 3

CONTRACTING APPLICATIONS

3-1 USING PRICE INDEX NUMBERS IN DEFENSE CONTRACT PRICING*

The contracting community has need to employ many different tools and techniques in estimating the prices of future purchases. The high probability that costs and prices of goods and services will change over time requires development of approaches to estimate that change. One set of approaches uses index numbers to forecast the change in price of goods and services.

The price analyst uses index numbers for three general purposes: (1) to deflate or inflate prices for comparison analysis, (2) to project price or cost escalation in contractual documents, and (3) to inflate or deflate costs to facilitate trend analysis. Index numbers are used in price analysis to compare the proposed cost of an item with the cost of the same or similar item procured in past years. Here, the index numbers are used to discount inflation that has occurred over time so the comparison can be made in constant year dollars. Escalation clauses usually call for some kind of after-the-fact pricing action adjusting the price paid to reflect actual price levels at the time of contract performance. These clauses use index numbers to measure the change in price levels over time. Index numbers are also used to facilitate trend or time-series analysis of individual cost elements by eliminating or reducing the effects of inflation. The analysis can then be performed in constant dollars. Each of these three uses of index numbers in contract pricing is discussed along with some examples.

Price Comparison Analysis

One of the uses of price index numbers is to measure price inflation. One can define price inflation as the time related increase in price of an item or service of constant quality and quantity. Price index numbers can be used to compare the prices of the same or similar items purchased in different time periods by inflating the old purchase price to a current time period or deflating a current price to some old purchase time period.

Consider the problem of analyzing a contractor proposal of $85,500 for a turret lathe to be delivered in 1976. A procurement history file reveals that the same machine tool was purchased in 1972 at a price of $48,500. The task is to determine if the proposed price is fair and reasonable.

---

*Smith, Larry L., The Use of Index Numbers in Defense Contract Pricing, Technical Report, AU-AFIT-SL-1-76, November 1976.

The approach is to: select or construct an appropriate index series, forecast the series to the anticipated date of production, inflate the old price to current dollars and compare. The Machinery and Equipment Subindex of the Wholesale Price Industrial Commodities Index (BLS) is selected as a reasonable indicator of price movement for the item. The data of Table 3-1 are extracted from the 1976 issue of the "Economic Report of the President".

Table 3-1
MACHINERY & EQUIPMENT

Year	Index	Year	Index
1967 (base)	100.0	1973	121.7
1971	115.5	1974	139.4
1972	117.9	1975	161.4

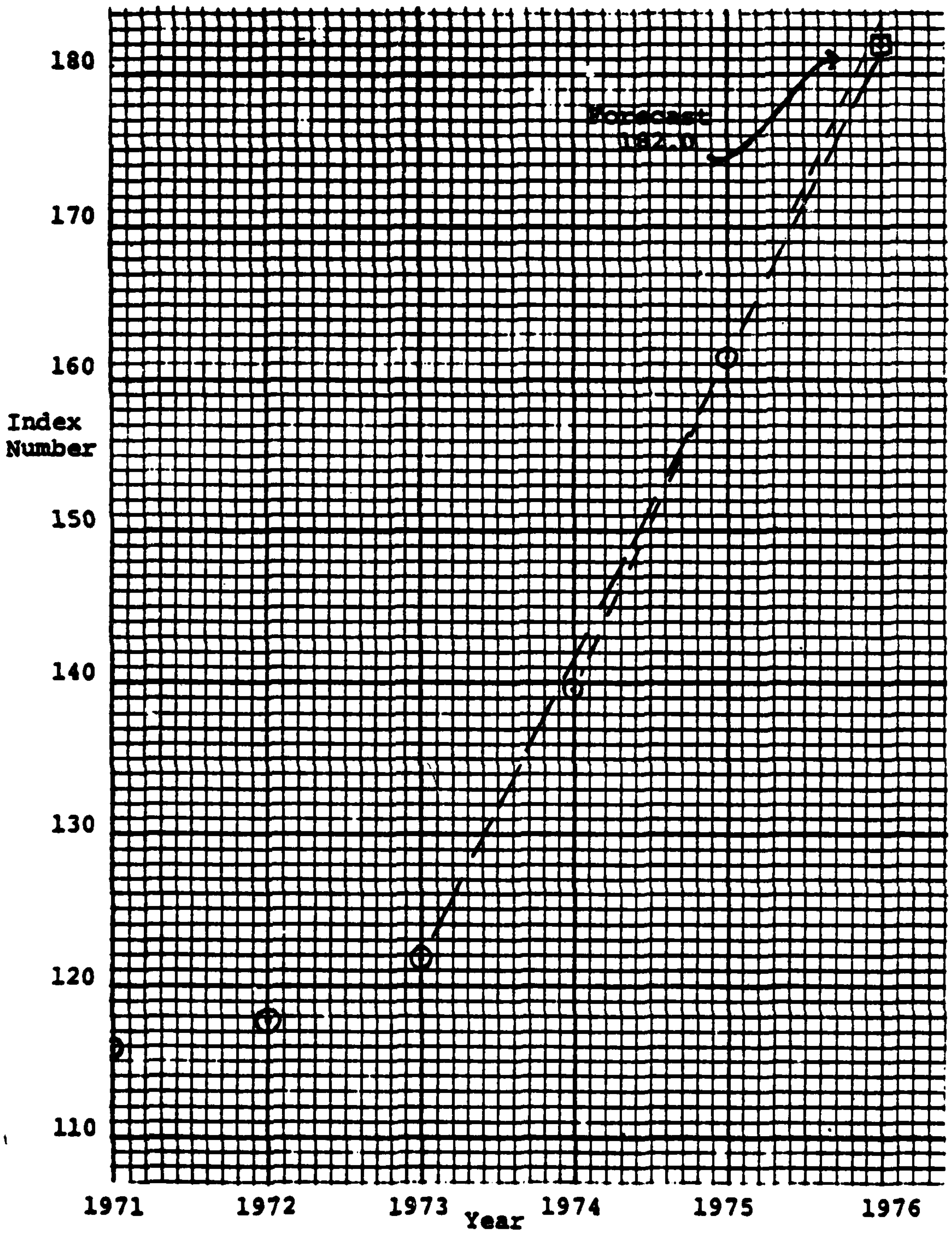
The data need to be forecasted to cover the anticipated period of production; 1976. This can be done by first graphing the data on rectangular coordinates, hand fitting a curve to the data, and projecting the curve into the future. See Figure 3-1.

For short range forecasts (2 years or less) the most recent data usually give the best indication; therefore, the last two data points are used for a straight line projection to 183.0. A line drawn from the third from last point through the last point gives a check of the forecast at 181.0. Thus an index number for 1976 between 181.0 and 183.0 appears reasonable. Third, inflate the 1972 actual price to 1976 dollars and compare the result with the proposed price of $85,500. This is done by first deflating the 1972 price to 1967 dollars ($48,500 ÷ 1.179 = $41,137) and then inflating this result to 1976 dollars ($41,137 X 1.820 = $74,869). This $74,869 is in 1976 dollars and gives one indication that the proposed price of $85,500 is excessive.

A few comments are in order here to highlight possible sources of error. First, the approach assumes constant quality and quantity. That is, the procuring agency is buying essentially the same item in essentially the same quantities as purchased in 1972. Second, it is assumed that the general Machinery and Equipment Index is representative of a specific company's turret lathe. In fact, the Machinery and Equipment Wholesale Price Index is made up of samples of production of many different kinds of machines and equipment produced by different contractors located all over the country. Third, it is assumed that the past will forecast the future and that a line drawn through data points will predict future inflation. This might be wrong as few analysts can forecast an economic turning point.

Nevertheless, the index number approach to price analysis gives the analyst another tool for comparison. It can be used to check prices predicted by other methods of analysis such as parametric or detailed analysis approaches. The index number approach can also be used as a basis for price negotiation when the buyer is lacking substantive price data.

Figure 3-1
Forecasting an Index Number



Price Escalation Clauses

There is a need for some Government contracts to contain a price readjustment arrangement providing for significant unanticipated fluctuations in the economy. This need is most apparent in those contracts that call for performance a long time in the future (more than two years), although sometimes the need exists for shorter term adjustment. Contracts that fall in the longer range category are typically large systems production contracts or multiyear production contracts.

The possibility of unanticipated fluctuation in the economy is one of the elements of cost risk in any contract. As the period of performance becomes further removed from the time the contract is written, the risk becomes greater. Contractors normally include some contingency dollars in a cost proposal to compensate for this risk of cost overrun. As the risk becomes greater with longer periods of contractual coverage, the price of the contingency becomes unacceptably high from the Government's point of view. This risk can be shifted in part or in total to the Government through the use of a price adjustment clause.

One can identify two general approaches to constructing price adjustment clauses for economic fluctuations in large dollar, extended production efforts. One is to construct a clause to compensate for any and all changes in the economy. The other is to construct a clause to compensate only for abnormal fluctuations in the economy. Both approaches use some index number series as a basis for making a price redetermination at the conclusion of the effort.

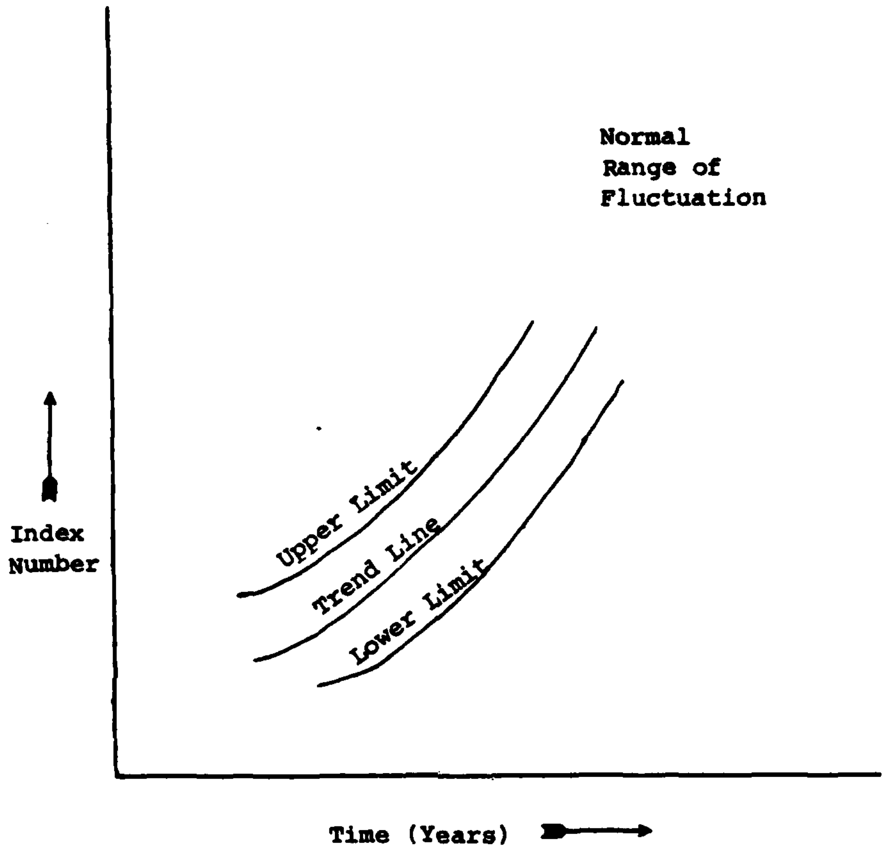
The "any and all" fluctuations approach is simpler and easier to understand. If the index number forecast for some specified future period of performance is higher or lower than actual, then the contract costs originally predicted using the forecast index number would be adjusted to reflect actual index numbers extant at the time of performance. For a specific element of cost, the "any and all" fluctuation clause might include a repricing formula as follows:

$$\text{Price Adjustment} = \frac{\text{Actual Index} - \text{Forecast Index}}{\text{Forecast Index}} \times \text{Target Cost}$$

The probability of correctly predicting the level of the economy and its associated index number is low, thus making the need for some repricing highly probable. If the index numbers chosen accurately relate a particlar product to the economy, this approach eliminates all of the contractor's risk associated with fluctuations in the economy.

The "abnormal" fluctuations approach assumes that a normal range of economic change can be defined and that prices will be adjusted only if the economic indicators fall outside of that normal range. Figure 3-2 portrays the concept.

Figure 3-2
Fluctuation Around an Index Number Trend



This approach requires two formulations of the adjustment equation for each element of cost to be repriced. One formulation adjusts the costs upward if the actual index exceeds the high side of the range and the other formulation adjusts the costs downward if actual index falls below the low side of the range. For example:

$$\text{Adjustment} = \frac{\text{Actual Index-High Forecast Index}}{\text{High Forecast Index}} \times \begin{array}{c}\text{Target}\\ \text{Cost}\end{array}$$

$$\text{Adjustment} = \frac{\text{Actual Index-Low Forecast Index}}{\text{Low Forecast Index}} \times \begin{array}{c}\text{Target}\\ \text{Cost}\end{array}$$

Selection of the range itself is a rather arbitrary process, but it can be generalized that the wider the range, the greater the amount of risk shifted to the contractor. A reasonable range might be plus or minus one standard deviation of the data from which the trend line was originally defined. Figure 3-2 is a graphical portrayal of the concept. In addition to selecting the general approach to construction of the clause, "all or any" versus "abnormal," the contracting parties must agree on some other variables in selecting methods for price adjustment due to economic changes. Two of these other decisions are whether to apply the adjustment to forecast costs or actual costs and whether to use the forecast index or the actual index as the denominator in the adjusting fraction.

The adjustment might be a function of either target costs or actual costs, e.g.:

$$\text{Adjustment} = \frac{\text{Actual Index-Forecast Index}}{\text{Forecast Index}} \times \text{Forecast Cost}$$

$$\text{Adjustment} = \frac{\text{Actual Index-Forecast Index}}{\text{Forecast Index}} \times \text{Actual Cost}$$

Opponents of actual costs as a basis for adjustment advance the thought that if overruns or underruns occur, the amounts of the overruns or underruns would unfairly influence the adjustment. Another argument against the use of actual costs is the difficulty in defining them since allowability of costs is often subject to negotiation. A third argument against the use of actual costs as the basis for adjustment is time delay. Often actual costs are not audited for years after contract completion.

Advocates of actual costs as a basis for adjustment believe the purpose of the adjustment clause is to adjust the price for causes beyond the contractor's control. Therefore, the adjustment should be made on the basis of actual costs whatever they might be.

Another decision facing the constructor of escalation adjustment clauses is the need to decide on which factor to use in the denominator of the adjustment fraction. Precedent has established that the denominator be the actual index when actual costs are adjusted and that the forecast index be the denominator when forecast costs are to be modified, e.g.:

$$\text{Adjustment} = \frac{\text{Actual Index-Forecast Index}}{\text{Forecast Index}} \times \text{Forecast Cost}$$

$$\text{Adjustment} = \frac{\text{Actual Index-Forecast Index}}{\text{Actual Index}} \times \text{Actual Cost}$$

It is apparent that in a period of rising economic indicators, that the cost adjustment fraction (and the resulting adjustment) would be smaller if the actual index were used as the denominator. It is intuitively better to use such a smaller adjustment fraction with actual costs that already include the effects of inflation. On the other hand, for those adjustment equations based on forecast costs, it appears more fair to use the forecast index as the denominator of the adjustment fraction.

A third decision facing the constructor of escalation clauses is the choice of index series and the weights to apply to each series used. For example, one Air Force systems acquisition contract for aircraft uses two economic indicator series to adjust the price for inflation. The two series are the BLS, Wholesale Price, Industrial Commodities Index and a wage rate series from the BLS yearly average wage rates for the Durable Goods category of Production Workers in Manufacturing by Major Manufacturing Groups. The Industrials Commodities Index is chosen to represent the materials portion of the aircraft price and is weighted 25% of the total price. The wage rate series is chosen to represent the labor portion of the price and is weighted 75% of the total price. The redetermination clause is of the form:

$$\text{Adjustment} = \frac{\text{Actual Index-Forecast Index}}{\text{Forecast Index}} \times \begin{matrix}\text{Forecast}\\ \text{Cost}\end{matrix}$$

The clause constructor needs to trade off clause complexity and the subsequent difficulty of constructing and administering the clause against level of accuracy desired. The above mentioned aircraft procurement escalation clause is simple and easy to construct and administer. The contracting parties take the risk, however, that the economic indicator series are not closely related to the airframe manufacturing industry. There is little consideration of indirect cost factors which often change at a different rate than direct cost factors.

The method chosen to forecast the index trend is a fourth major consideration in clause construction. A straight line regression analysis built on price index numbers of the 1960s will seriously underestimate price indexes of the 1970s. It is a good idea to plot the data, discern a trend and best fit a trend model through statistical analysis for projection purposes. A straight line has not been a good tool for such projection in recent years.

The ideal contract clause for adjustment of price due to inflation does not seem to exist. If the clause closely approximates the economic changes of a specific product, it is complicated to construct and administer. If the clause is simple to construct and administer, it does not relate well to the product for which the economic adjustment is sought. Recognizing that all solutions are compromise solutions, the following ideas are suggested as a starting place from which to construct an adjustment clause.

First, select the normalcy band adjustment clause form. For example:

$$\text{High Side Adjustment} = \frac{\text{Actual Index-High Side Forecast Index}}{\text{High Side Forecast Index}} \times \text{Forecast Cost}$$

This adjustment form uses a forecast index and a forecast cost (target cost). By using a forecast cost as a basis for adjustment, the administrative problems associated with determining allowability of actual costs are avoided.

Second, select at least three index series to represent the price of the product. One series should be related to the materials used in or purchased for the product being priced. The second series should relate to direct labor of the type used to produce the product. The third series should relate to the indirect costs on the product being priced. Weights totaling 100% need to be assigned to each category of costs. After selecting the index series, each needs to be forecast for the period of performance. For short range forecasts (less than 2 years) an exponential smoothing model best fitting each set of data should be used. For long range forecasts (2 - 10 years) a trend model best fitting each set of data should be used. Generally speaking, one should use more than 10 years of data to discern this trend.

Third, the forecasted index numbers should be used to inflate the out-year price estimate from current dollars to future dollars for target pricing purposes. This task is necessary whether or not an economic adjustment clause is included in the contract.

Fourth, a width of the normalcy band for each series of data must be defined. A zero width band means that the Government assumes all the risk of inflation. Rather arbitrarily, a normalcy band of ± one standard error of the mean is suggested as a negotiation starting point for a normalcy band. This affords the contractor protection against "abnormal" inflation and the Government still does not assume all the risk. Each economic series would have a normalcy band. If actual index numbers move outside the normalcy band for the period being priced, the contract should provide that the price of the affected element (materials, labor or indirect) would be adjusted in accordance with the formulation. Such a clause should contain as a minimum; (1) the formulation of the repricing arrangement, (2) the index series source to be used to determine the actual indexes when repricing, (3) an example of how the two parties expect the repricing formula to work if it is exercised.

Price Deflators to Facilitate Regression Analysis

Price deflators can be used to deflate prices to a constant dollar basis to facilitate regression analysis. Many times cost data are aggregated by function or category of incurrence for accounting purposes. These data are also related to the flow of time because of the practice of keeping accounts by accounting period. An example of such an account would be a general and administrative (G&A) account, one of the contractors indirect cost pools. G&A pool data are usually available on a yearly basis

as far back as the contractor keeps records. When the analyst needs to predict costs such as G&A costs for future years of operation, he needs to find some level of activity indicator that causes G&A costs to vary. An example of such an indicator might be the number of direct employees, number of total employees, cost of sales or other similar independent variables.

Before one begins the regression analysis of the dependent variable (pool dollars) on the independent variable(s) (level of activity indicator), the dollars need to be deflated to a constant dollar basis. This is because the dollars were actual dollars expended in different periods of time. These dollars include the effects of inflation and they have different values.

By selecting a price index number series that represents the kinds of costs that are in the aggregation, and deflating each dollar grouping by its corresponding index number, the effects of inflation can be removed. The dollars are converted to a constant (base year) basis suitable for regression analysis, forecasting and subsequent reinflation for pricing purposes. The following example is intended to illustrate the use of index numbers when forecasting manufacturing overhead costs for a future year of contractor operations. Table 3-2 is a summary of cost data keyed to the explanatory notes that follow.

Table 3-2
COST DATA SUMMARY

Year Number	1	2	3	4	5	6
Year	1967	1968	1969	1970	1971	1972
Actual Cost[a]	$32,670	$41,111	$43,879	$51,276	$59,432	
Activity Level Direct Labor Hours (DLH)[b]	17,800	21,000	22,800	27,100	35,900	
Index Numbers[c]	100.0	104.8	108.9	112.4	114.8	
Deflated Costs[d]	$32,670	$39,228	$40,293	$45,619	$51,170	
Activity Level Forecast (DLH)[e]						35,000
Deflated Overhead Cost Model & Deflated Forecast[f]	Deflated OH \$ Cost = A + B (DLH) = 17,688 + 0.96734 (DLH)					51,544
Index Nr. Cost Model & Forecast[g]	Index No. = A + B (year number) = 97 + 3.72 (year number)					119.3
Actual Cost Forecast[h]						61,492

TABLE 2 - Continued NOTES

[a]Actual costs for past years of operation are collected from the contractor's accounting records. These costs may need to be adjusted to compensate for changes in the accounting system from year to year.

[b]The activity level selected to explain the variation in overhead costs from year to year is direct labor hours (DLH). Historical DLHs are collected for years corresponding to the overhead cost data.

[c]The index numbers listed here were constructed from contractor's indirect cost data for prior years using the weighted average of Price Relatives (current year weights) approach. The analyst might also have selected a previously constructed index series that was representative of the contractor's overhead cost pool make-up.

[d]The actual costs are deflated to constant 1967 dollars by dividing each year's overhead costs by the corresponding index number.

[e]The forecast of direct labor hours for the future year(s) is a function of the contractor's sales projection. Necessarily then, the contractor must assist the analyst with the forecast.

[f]The deflated overhead cost model is a simple straight line derived by regressing deflated costs on the corresponding activity level (DLH). The deflated overhead cost forecast is calculated from the cost model by inputting the forecast direct labor hours, 35,000.

[g]The index number cost model is a straight line derived by regressing the index number of the corresponding year number (1,2,3,4 and 5). The index number forecast was calculated from the cost model by inputting year 6.

[h]The actual cost forecast is calculated by multiplying (inflating) the forecast deflated overhead costs 51,544 by the forecast index number, 119.3.

3-2 TIME-SHARING SERVICES AND COPPER IMPACT*

Computers and Pricing

By its very nature, the contract pricing and financial analysis process involves two major factors: first, the informed judgment of the analyst based on experience and training; and second, a considerable amount of computation. A comprehensive study of the pricing function in 1972 revealed that a disproportionate amount of the professional price analyst's time was being spent in computations and related mechanical and administrative tasks, leaving little or no time for the analyst to thoughtfully probe all issues in a cost or price proposal. The principal computational tool was the standard mechanical calculator.

*Sandman, Tom. Contracting Applications of Copper Impact, Department of Contracting Management.

The obvious candidate to assist the analyst with his computation task was the digital computer. This offers several features of interest to the analyst:

- Accuracy; the computer in its present state rarely makes computational errors which are not based on improper programming or faulty data.

- Speed; once a job enters the computer, it is capable of easily doing computations in a few seconds which would take hours manually.

- Flexibility; once a program is established, any number of data sets may be processed in an identical manner or, conversely, with a minor program change, the same data set can be processed differently.

Until the late 1960's, bringing the power of the computer to the field analyst was difficult and time-consuming since it usually involved use of someone else's program and passing information and data through the mail to the computer center. The advent of the time-sharing computer solved this problem by providing a means to gain on-demand access to the computer central processor from a remote terminal over telephonic communication lines.

Overview of Time-Sharing

The time-sharing computer is a complex of devices which organize and direct multiple access to a single central processing unit. A job control computer processes several jobs simultaneously in the same central processor by efficiently utilizing the computation space of the central processor. Through a priority system and a complex set of queuing logic, individual steps of several jobs can be organized in the central processor so as to make the most efficient use of the time available.

The other half of the time-sharing equation relates to communication with the central processor. Through the use of communications terminals (teletype or CRT) and a network of communication lines, the user of the computer can be located virtually an unlimited distance from the computer itself. From the user's standpoint, however, the computer might as well be in his office since all the computational power is delivered to him via the terminal and a set of very specific words called system commands.

Overview of Applications

The focus of this section is to catalog and describe the myriad of applications of the time-sharing computer to the pricing and financial management task in the Air Force. The applications generally fall into five major categories, discussed in the following paragraphs:

a. Cost Models

Cost proposal logic simulation is a more appropriate phrase for the process of simulating with a computer program the logical build-up and cost relationships of a specific cost proposal. Given a model of a contractor proposal, the analyst can very quickly:

- Arithmetically audit a proposal.

- Compute his own negotiation objective based on the results of his analysis.

- Recompute his position during negotiations.

- Generate neatly printed and formatted outputs of the traditional "spread-sheet" display of cost proposal analysis and results (usually in tabular form).

Another reason for cost models is the performance of "sensitivity" analysis, or the "what-if" game. The analyst, through a series of model runs, can determine which basic cost elements drive the most significant changes in bottom-line price features.

b. Overhead Management

Several years ago, the Air Force developed PIECOST, which was an early initiative in automated overhead negotiation and tracking procedures. Basically, PIECOST extrapolated prospective pricing data from history using various statistical techniques. Since those earlier days, computer applications in the overhead area have evolved into sophisticated cost accounting system simulation models. The old PIECOST acronym has been replaced by "MODE", referring to Management of Overhead Discrete Evaluation. In essence, these models provide the Air Force Plant Representative Offices with greater visibility into indirect cost build-up from incurrence to final rate development. These models enhance the negotiation of forward pricing rates by performing the time-consuming manipulation of discrete cost inputs. The advantages of such models are essentially those of the cost models discussed previously.

c. Workload Management

This application area is largely self-explanatory. The basic approach is to automate the pricing case register in such a way as to satisfy the following information needs:

- Local management visibility of case load to facilitate resource allocation decisions.

- Internal management information for overall management decisions.

- Higher headquarters workload information for policy decisions and surveys.

d. Data Banks

The computer has long acted as an excellent medium for the storing and retrieval of information. The time-sharing computer permits a wide range of possibilities for centrally storing and maintaining information to

be used at widely dispersed locations. The primary application in COPPER IMPACT is the centralized pricing rate and factor data banks, CONRATES, discussed elsewhere in this paper.

e. Analysis Aids

Numerous individual programs have been developed under COPPER IMPACT whose sole purpose is to aid the price/financial analyst in performing in-depth analyses. In addition, the contractor's software library contains numerous programs of interest to pricing and finance personnel. These programs are available to any user through the Copper Impact Library.

Copper Impact Applications

Introduction

A broad overview of the nature of computer applications to contract pricing was presented in the previous section. This section includes additional detail which will form the foundation upon which a user may build his individual application.

This section is not a user guide to each program, but rather a reference guide to application concepts and known programs which have been developed to date. Further detail on the specifics of using programs in this section should be obtained from the responsible organization noted with each given program and/or system.

Cost Proposal Simulation Models

a. Overview of Modeling

The purpose of a cost proposal simulation model is to provide the price analyst with a computational tool of considerable worth in reducing the mechanical tasks related to cost/price analysis. Such a model provides a tool to audit a proposal, develop a position, recompute a position, or conduct sensitivity analysis. These tasks are accomplished quickly and accurately by the computer which also generates a neatly typed output tabulation suitable for direct inclusion in a formal report or case file without being retyped.

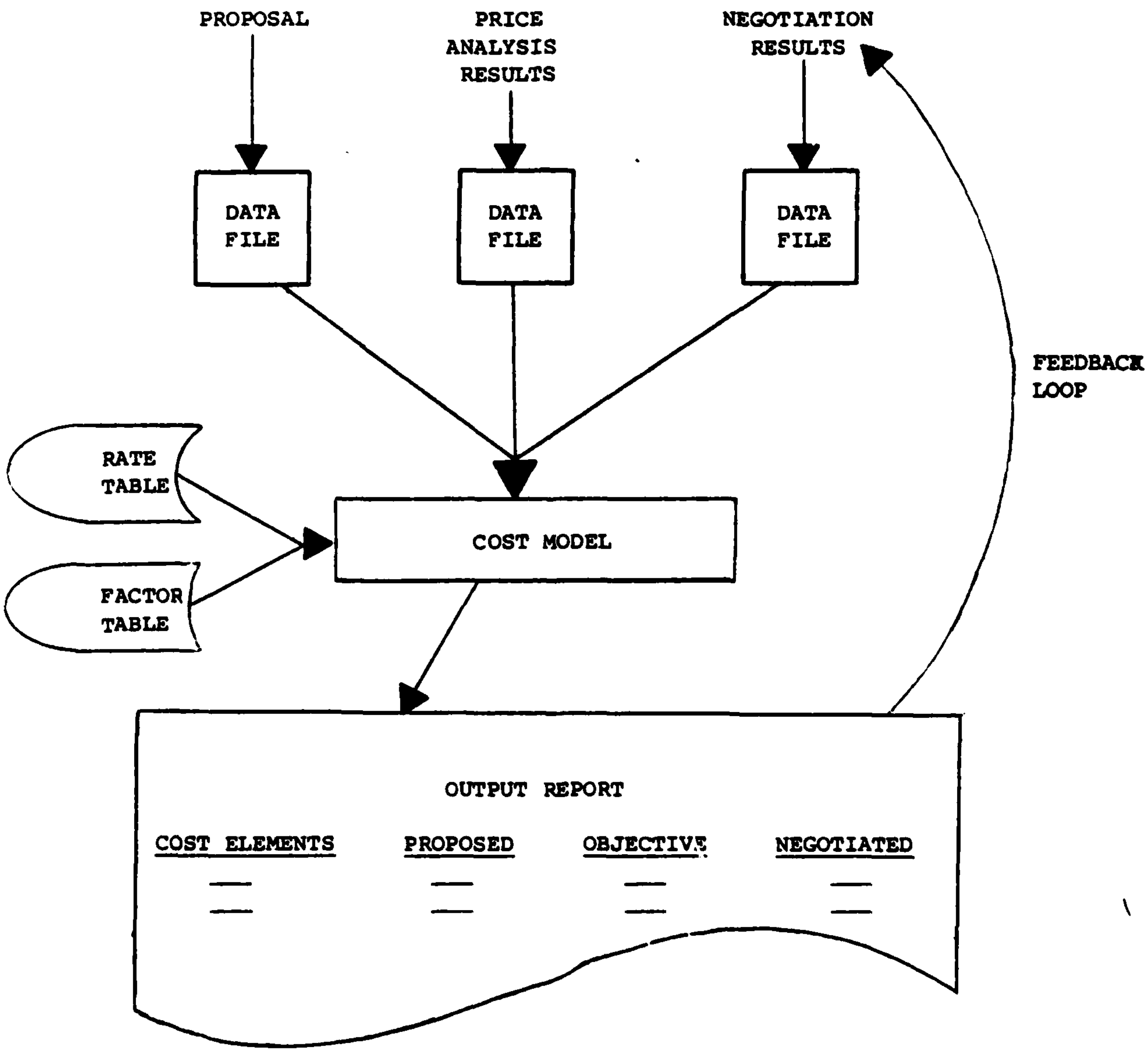
Cost models are built for a specific type of proposal format from a given contractor. The model itself is structured to accurately reflect the arithmetic relationships of basic inputs (direct cost amounts, rates, and factors) and their build-up to the final proposal amount. The model usually provides for an output report which displays the three standard elements, proposed, objective and negotiated. Figure 3-3 shows a typical model structure and flow of information.

Developing a cost proposal model poses several questions which need answers prior to beginning the actual task of programming the model. The type of considerations to be made include the following:

- A complete, detailed understanding and working knowledge of the contractor's estimating and proposal system is essential. An important element is the identification of those costs which must be direct inputs versus those which are derived from rate or factor application.

- Are the contractor's proposals sufficiently similar in format and logic that one or two models can be developed which handle all proposal types? Alternatively, must a separate model be constructed for each proposal type?

Figure 3-3
COST MODEL STRUCTURE



- Are rate and factor structures sufficiently general that a single rate table can be used for several different proposal types?

- Is a single output report format satisfactory or are multiple report options desirable? Are time-period spreads of data necessary?

Cost models have proven very useful in all aspects of the pricing process, and savings have been documented at from 3:1 to 10:1 in terms of manhours consumed in pricing computations and report generation.

b. Proposal Pricing System (PPS)

COPPER IMPACT has sponsored the development of two cost model building systems which permit the user without knowledge of a programming language to construct cost models for specific proposal types.

One of these systems is called the Proposal Pricing System (PPS). The system is fully described in the PPS User Guide available from Headquarters AFSC/PMMP, Andrews AFB, DC 20334. The system relies on a set of codes as inputs to describe the model for a model building program. Input data for a given proposal is input from a second file, processed, and the results put in a master data file. Reports are generated with either the standard system report modules or specialized report routines prepared by the user. Figure 3-4 provides an overview of the flow of information in the PPS System.

The principal utility of PPS is in reducing the amount of time required to develop a cost model using a programming language such as BASIC or FORTRAN. Savings in this area range from 6:1 to 10:1. These savings are, however, offset by the cost of running each case through the model. Therefore, as a rule, if a model is to be used frequently throughout the year, the model should be specifically programmed in a language (BASIC or FORTRAN) rather than using PPS.

c. Programmable Cost Model

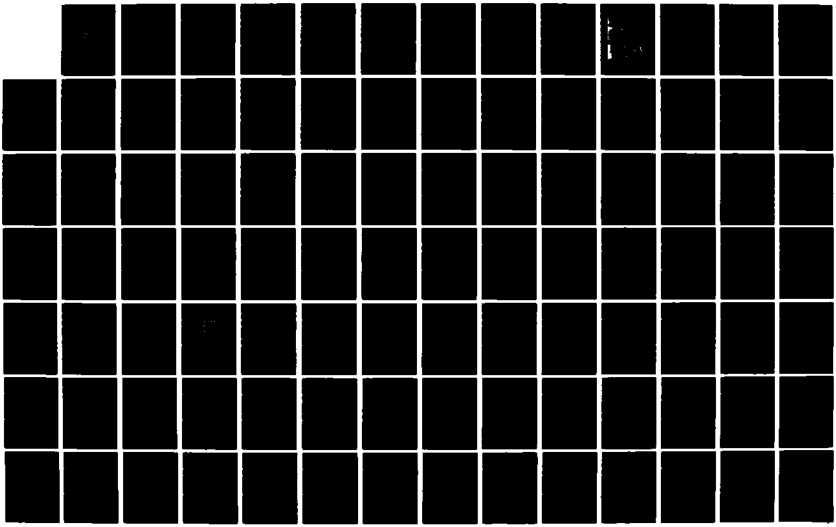
The Programmable Cost Model is identical in purpose to the Proposal Pricing System, but relies primarily on an approach in which the operator interacts conversationally with the program when defining the model and providing data input. Specific information and detailed user documentation can be obtained through Headquarters AFLC/PPPP, Wright-Patterson AFB, Ohio 45433.

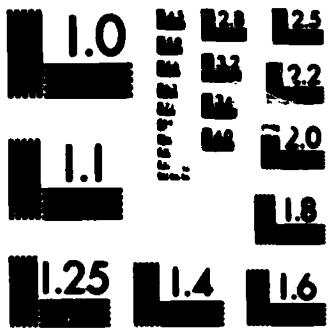
d Specialized Cost Models

Many specialized cost models have been programmed by COPPER IMPACT users to apply to specific cost proposal formats from specific contractors. These models are available to any user through the COPPER IMPACT Library.

AD-A121 990 QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS J E ENGEL APR 80 AFIT-LS-32 2/7

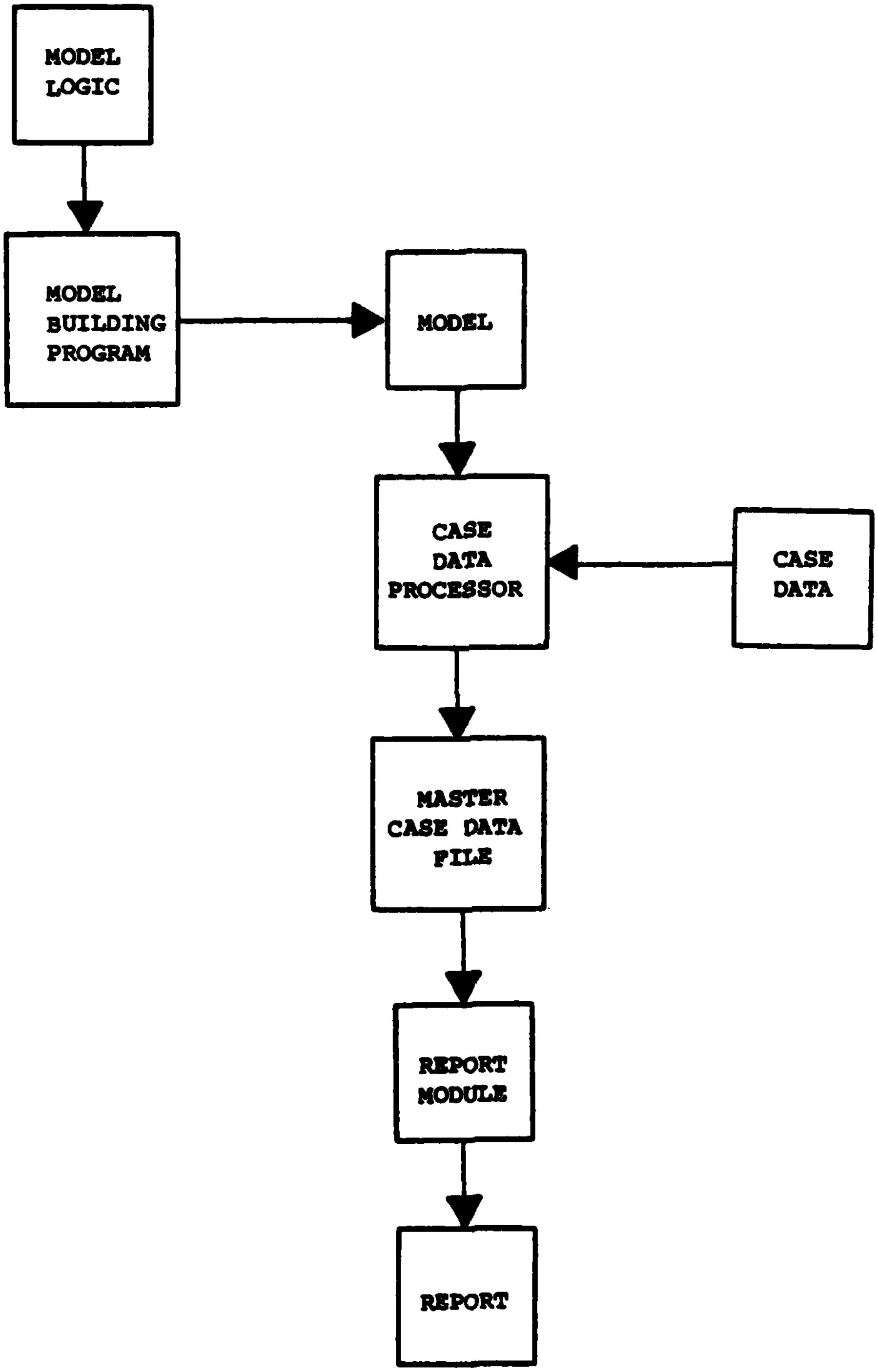
UNCLASSIFIED F/G 15/5 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Figure 3-4
PPS DATA FLOW



e. <u>Financial Analysis Language (General Electric FALII)</u>

FALII is a contractor-provided software package designed to provide tabular reports of a financial analysis nature. The language uses simple English-like statements and does not require a high degree of expertise in programming in order to be used effectively. By using a series of rows and columns, the user can establish a program which processes data in numerous ways. The system has been used effectively in establishing cost proposal models, but has numerous other applications such as time-phased cash flow analysis and budget development. As with most other software packages of this type, the overhead involved in making the system easy to use and quick to implement also makes it somewhat expensive. A user should be prepared to trade operating cost for development time in using FALII.

<u>Overhead Management</u>

a. <u>Overview</u>

Since the late 1960's, the DOD procurement community has been placing increasing emphasis on the management and monitorship of indirect cost. Organizations devoted to overhead management have been established which utilize advanced computer applications to enhance their cost analysis and negotiation functions. A prime example of this type of organization is the Business Management Branch within an Air Force Plant Representative Office (AFPRO).

b. <u>PIECOST</u>

One of the original applications of the time-sharing computer to the overhead cost analysis process was an approach known as PIECOST (Probability of Incurring Estimated Cost). The basic premise of PIECOST was that indirect expense, when properly categorized, could be forecasted on the basis of historical extrapolation of pricing data through the use of correlation and regression analysis. Experience has shown that while the concept may be valid, the forecasting technique and indeed the bulk of the approach proved to be of limited use in negotiating actual rate agreements with contractors. This is not to say that statistical techniques are inappropriate in the contract pricing function. On the contrary, they are used by the Air Force when it is practical to do so.

The primary benefit of PIECOST was that it provided a philosophical framework from which the current computer applications were developed. Some of the computer programs used within the PIECOST application package are still used as cost tracking aids. Specific details on their use can be obtained from the Air Force Contract Management Division (AFCMD/TMO, Kirtland AFB, NM 87117).

c. <u>MODE (Management of Overhead Discrete Evaluation)</u>

The MODE model is the principal computer aid used by the AFPRO overhead cost analyst. Each MODE model is tailored to simulate a

particular contractor's cost accounting system. MODE provides greater visibility of overhead cost flow from incurrence to final rate development. These models have been implemented at most AFPRO locations where they are used primarily for negotiation of Forward Pricing Rate Agreements or Recommendations (FPRA or FPRR). Detailed information concerning the implementation and use of these models can be obtained from AFCMD/TMO.

d. SORT (Settled Overhead Retrieval Technique)

The continued concern at OSD and Service level about consistency in the way overhead is managed prompted AFCMD/TMO to develop a computerized technique to maintain a track on the treatment of overhead cost items in final settlement. The purpose of SORT is to provide management and the ACO with data to show how overhead costs are being treated by various AFPRO locations.

When fully implemented, the SORT data base will contain final settlement findings for each AFCMD location for a three-year period. In brief, the validated user will be able to make inquiries of SORT on the basis of key word descriptors or DAR cost principles. These key words or DAR principles will relate to various overhead final settlement findings. SORT will respond by providing a printout of information pertaining to each finding relative to the particular inquiry.

Management Information Systems

a. Overview of Management Information Systems

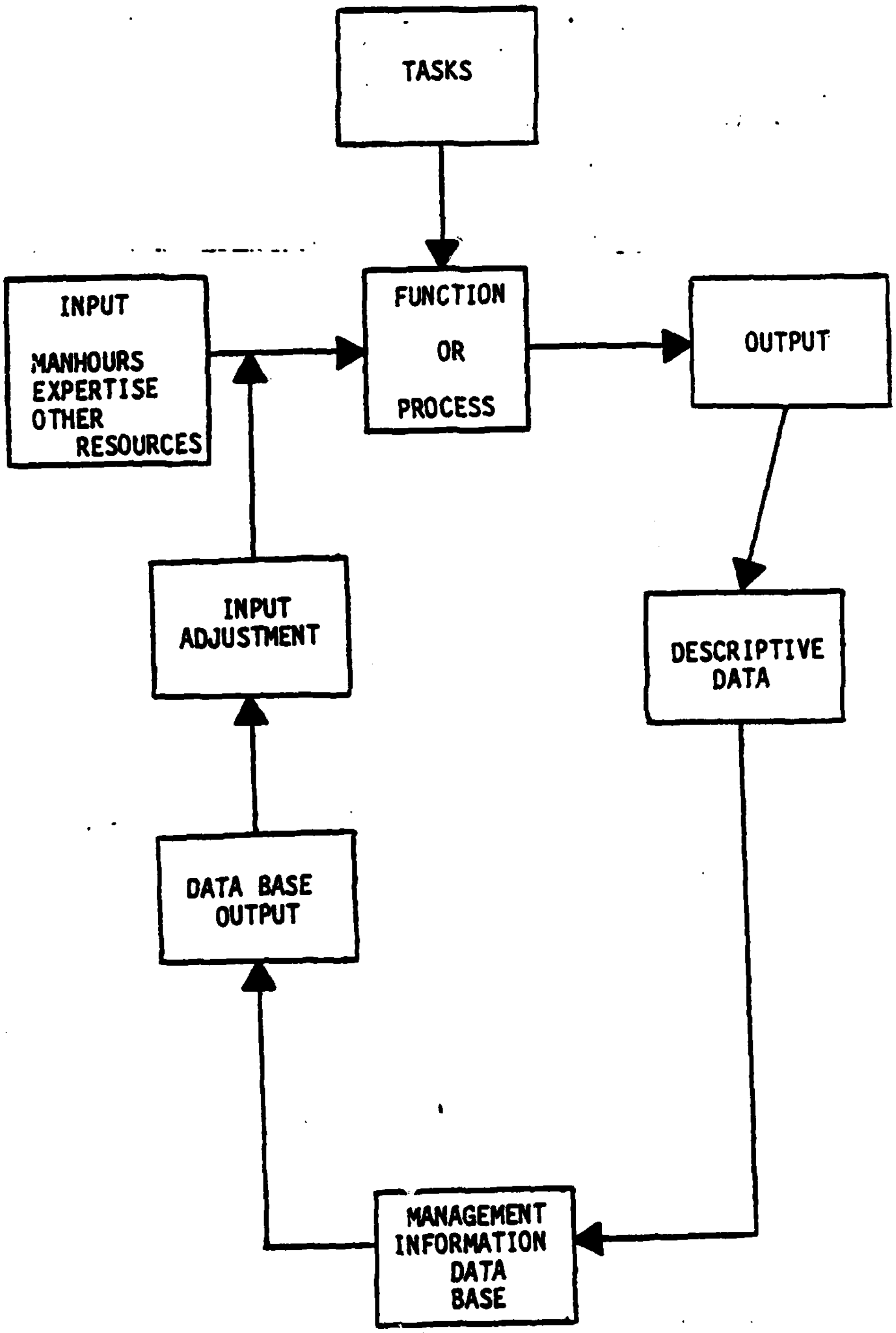
The advent of the computer has made management and detailed analysis of large amounts of data more feasible and, in turn, more fruitful. A properly organized automated data base is in many cases the only way to convert otherwise massive amounts of data into valuable management information.

b. AFCMD Pricing Management Information System

An information data base in terms of workload management is the key element in the feedback loop to management. By structuring inputs and storage rules based output needed by management, an eventually indispensable tool can be created. Figure 3-5 is a schematic diagram of a typical management control system which incorporates an automated feedback loop.

Until 1976, automated workload management systems under COPPER IMPACT were developed purely on a local basis by individual user organizations. These systems utilize a variety of approaches to the problems of output requirements, required inputs and information processing methodology. Based on a discerned need at the headquarters level for enhanced pricing workload data, the AFCMD Pricing Division developed a centralized pricing workload data base in 1975. This Pricing Management Information System (PMIS) will serve the information needs of both field pricing chiefs

Figure 3-5
Typical Management Control System



in local resource management and headquarters personnel in pricing policy development. The data base design anticipates application to both AFPRO and buying activity pricing offices, and incorporates optional data to record manhour data for each pricing case. Detailed information and user guides can be obtained from the Air Force Contract Management Division, Pricing Division (AFCMD/TMF, Kirtland AFB, NM 87117).

c. Contractor Rates System

Description. The Contractor Rates System (CONRATES) is an automated data bank designed to contain rates and factors used by price analysts and buyers in the task of forward pricing of contracts. The data bank was not intended to contain historical rates and/or factors.

CONRATES includes data on a contractor-by-contractor basis for each of up to five different categories: direct labor rates; overhead rates; direct cost factors; contract performance information; and factors used in pricing technical data. At the present time, emphasis is being placed on use of the first three sections of the data bank.

Each section of CONRATES for a particular contractor can contain up to five years of data in any of four forms: monthly, quarterly, semiannual, and annual. The system provides for two types of rates to be included; these are ACO recommended or negotiated. There is no practical limitation on the number of rate categories included in each section.

Responsibility for input and maintenance of rates and factors for a particular contractor rests with the cognizant Principal Administrative Contracting Officer (PACO). As a practical matter, data will only be input for the locations having a resident contract administration office. All such locations are encouraged to participate in the data bank, but emphasis is placed on those locations having a significant volume of pricing activity with more than one buying activity.

Access and Security. Access to CONRATES is restricted at several levels. First, the organization desiring data bank access must have or have access to a valid COPPER IMPACT user number/password combination. Second, the user must know the five-digit contractor identification code. Third, the user number of the organization must have been granted authorization by the PACO based on a justified need.

Where established communication channels are available, telephonic requests to the PACO are usually sufficient to gain the necessary approvals. In any case, a written request for approval should be provided to the PACO for his files and the user's own protection in the unlikely case of compromise of contractors' financial data.

Using CONRATES. From the user's standpoint, CONRATES is comparatively simple to use. Only two main programs are involved; the first, 'LISTCLC', generates an index to the particular contractor/section of index and the second, 'CONRATES', is the main data retrieval and maintenance program.

The primary identifier of data, after the contractor identification (CID), is the category/level code (CLC). 'LISTCLC' generates the numeric CLCs necessary to specify a report, the associated nomenclature, rate type, and period of data loaded. Once a user has established the CLCs of interest, 'CONRATES' is run in the REPORT or SELECT mode. The REPORT mode generates a neatly formatted report, complete with contractor identification data, containing the rates, application base definitions and any remarks applicable to each of the CLCs requested. The SELECT mode permits the user to specify a given CLC and time period(s) of interest which are returned without any special report format. A single rate for one time period may be retrieved in the SELECT mode.

Documentation. Contract administration organizations may request the CONRATES Input and Maintenance Guide from Headquarters AFCMD/TMF, Kirtland AFB, NM 87117. A companion User's Guide which provides instructions on accessing CONRATES may be obtained from Headquarters AFSC/PMMP, Andrews AFB, DC 20334.

d. Escalation Data Files

Two data files are maintained with monthly update on the COPPER IMPACT master library which contain information useful in surveillance of Economic Price Adjustment (EPA) changes. The files are described here for your reference. Additional indexes may be included on request to AFSC/PMMP, Andrews AFB, DC 20334.

Wholesale Price Index Data File (WPINDEX).

The WPINDEX file contains the following indexes by month for up to seven years of history:

(1) GNP Implicit Price Deflator

(2) Consumer Price Index

(3) Wholesale Price Index (All Commodities).

(4) Wholesale Price Index (Industrial Commodities)

(5) Selected detailed indexes based on those used in AFSC EPA clauses.

Hourly Earnings Rate Data (HERATES). The HERATES file includes the average gross hourly earning rates (including overtime) in the following categories (Standard Industrial Classifications) for up to seven years of history:

(1) Total private

(2) Durable Goods

(3) Selected SIC codes presently used in AFSC EPA clauses

e. Data Management System (DMS)

The Data Management System is a contractor-provided software package designed to provide a flexible tool for use by nonprogrammers in developing and managing a data base of any kind. The system employs a series of English-like statements which permit the user to establish and maintain a large data base. Reports are generated quickly and easily and may be developed for production use or one-time ad hoc inquiries to the data base.

Analysis Aids

Numerous individual programs have been developed under COPPER IMPACT whose sole purpose is to aid the price/financial analyst in performing in-depth analyses. In addition, the contractor's software library contains numerous programs of interest to pricing and finance personnel. These programs are available to any user through the Copper Impact Library.

3-3 USES OF THE LEARNING CURVE IN CONTRACT ADMINISTRATION*

Within the contract administration phase of the contracting process, there are several uses of the learning curve concept: (1) planning and scheduling of work to be accomplished by a contractor, (2) pricing of replenishment spare parts, (3) evaluating value engineering changes, and (4) pricing engineering change proposals. Each of these applications is discussed below.

Planning and Scheduling

The learning curve can approximate the number of man hours that will be required to produce a specific number of items. If the units are produced on a job lot basis, planners can also compute how many man hours should be scheduled for a given period. For example, a contractor experienced an 80 percent learning curve for an item he was producing with a first unit "cost" of 1,000 man hours. He provided the following production delivery schedule to the government indicating the man hours to be expended under the contract (Table 3-3).

Table 3-3
Contractor's Estimated Man-Hour Requirements and Delivery Schedule

Delivery date	Units to be delivered	Planned man-hour expenditure	Total cumulative manhours planned
3 June	10	6,315	6,315
17 June	10	4,170	10,485
1 July	10	3,000	13,485
15 July	10	2,800	16,285
29 July	10	2,500	18,785

*AFIT. Contract Administration. Volume II, Chapter 26, "Use of the Learning Curve in Contract Administration", 1975.

These data were plotted with a dotted line (scheduled man hours) on log-log paper in Figure 3-6. The analyst then plotted the cumulative average man hours required assuming an 80 percent learning curve. Examination of the data revealed that the planned man hours would not meet the delivery schedule required by the contract. The government's administrative contracting officer (ACO) was advised that the contractor would likely be five units short by the final delivery date unless there was a rescheduling of manpower. The analyst suggested to the ACO that the schedule reflected in Table 3-4 was more adequate in terms of manpower. Unless the contractor increased the planned man hour expenditure, he would be unable to meet his delivery schedule. The ACO informed the contractor of the discrepancy and requested a more realistic scheduling to prevent default of the contract.

Table 3-4
Government's Revised Estimated Man-hour Requirements and Delivery Schedule

Delivery date	Units to be delivered	Planned man-hour expenditure	Total cumulative manhours planned
3 June	10	6,315	6,315
17 June	10	4,170	10,485
1 July	10	3,535	14,020
15 July	10	3,173	17,193
29 July	10	2,929	20,122

Pricing Spare Parts

Direct labor is one of the major costs in most production contracts. The learning curve is used for the purpose of estimating a few high cost items produced over a long period of time and requiring a large amount of labor expended per unit. Suppose a contractor submitted the following unit price proposal to the ACO for the procurement of 50 units of a spare part:

Direct material (2,000 lbs @ $5.00 per lb)	$10,000
Direct labor (500 hrs @ $20 per hr).	10,000
Manufacturing overhead (@ 100%).	10,000
	30,000
General and Administrative expenses (@ 5%)	1,500
	31,500
Profit (@ 10%)	3,150
Unit Price Proposed	$34,650

Suppose that the contractor attempted to justify his proposal on the basis of an 80 percent learning curve. If so, the cumulative average labor requirement for 50 units would be 402.4 man hours. At $20 per man hour, the contractor should have quoted $8048 for direct labor. If all other factors remain the same, the contractor's proposal might be revised to reflect the following government negotiation position:

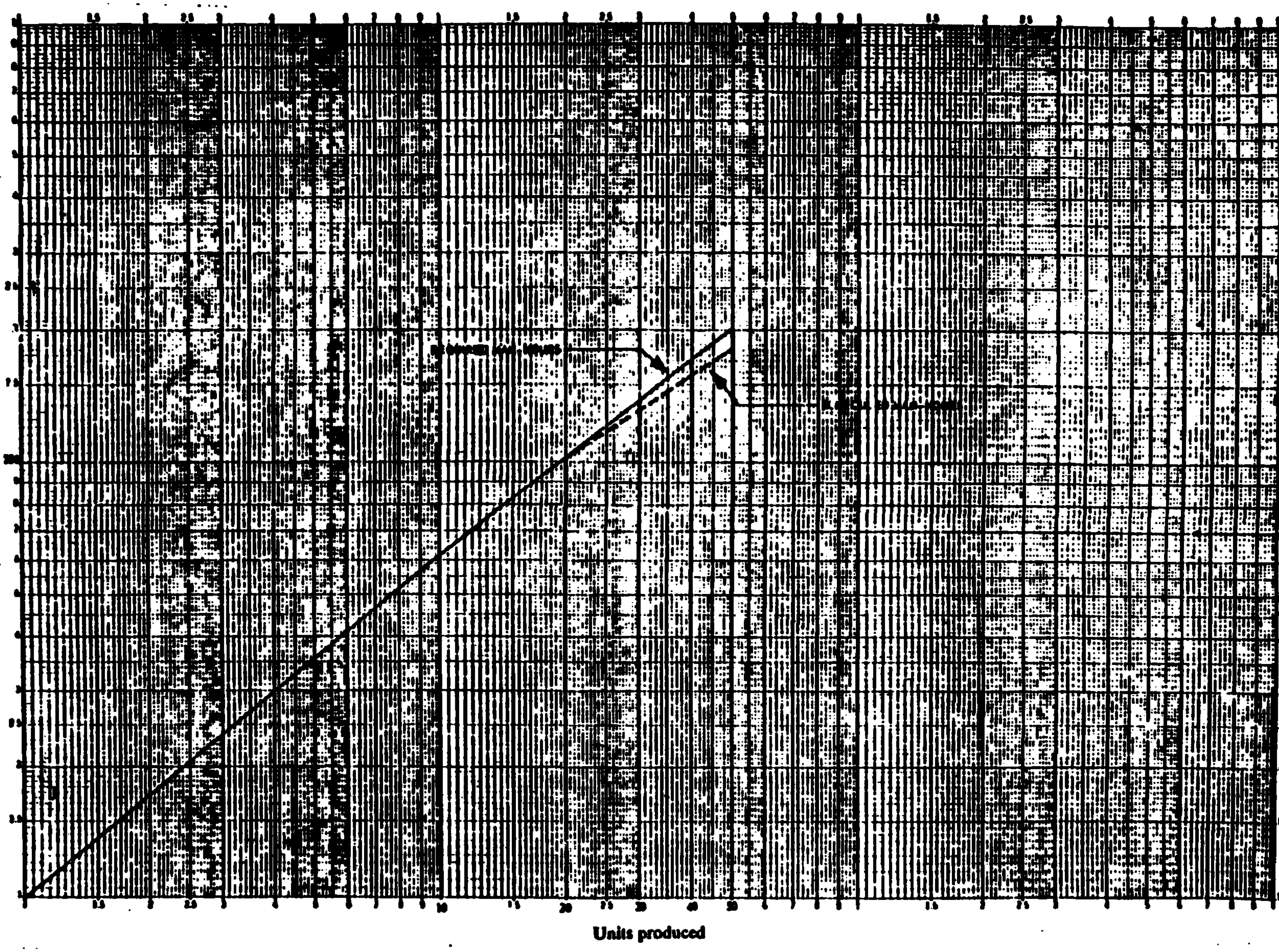


Figure 3-6 Comparing estimated man-hour requirements and delivery schedule.

Direct material (2,000 lbs @ $5.00 per lb)	$10,000
Direct labor (402.4 hrs @ $20 per hr).	8,048
Manufacturing overhead (@ 100%).	8,048
	26,096
General and Administrative expenses (@ 5%)	1,305
	27,401
Profit (@ 10%)	2,740
Unit Price Negotiation Objective	$30,141

Note the difference of $4,509 in the unit price proposed and the unit price negotiation objective suggested by the application of the learning curve. Whether the government can achieve its negotiation position would depend on the circumstances surrounding the procurement in question. The price proposal may be inflexible if the vendor knows that he is in a sole source position or that the matter is urgent to the government or that some other factor strengthens his bargaining position.

Evaluation of Value Engineering Changes

A value engineering change proposal (VECP) may result from a contractual provision which encourages the contractor to submit cost reduction ideas. Such proposals are reviewed by the ACO who makes a recommendation for approval or disapproval by the principal contracting officer (PCO). In submitting a VECP, the contractor must submit all necessary cost and technical data for evaluation by the government. Among the cost factors to be examined is direct manufacturing labor. When circumstances are appropriate and the necessary prerequisites are met, the contractor may estimate the man hours required based on the learning curve. A difference in the degree of slope may make one part of an item, which is proposed to be substituted for another, more economical and therefore more desirable in considering a VECP. The following example shows how the learning curve can be used in the decision to accept a VECP.

The contractor proposed to substitute part B for part A in an end item. Because part B is described by a 70 percent learning curve and part A by an 82 percent curve, the contractor claims that part B can be produced less expensively than part A even though the cost of producing the first unit would be much higher. Suppose the first unit of part A costs $350, and the first unit of B costs $500 (see Figure 3-7). The costs at unit 12 would be the same for units A and B, but for successive units, part B would cost less. If the government planned to buy 400 units, the 400th unit of part A would cost $63 in comparison to $23 for part B. In terms of total costs, the difference is even greater. For a procurement of 400 units, part A would cost $16,883 more than part B. Assuming technical acceptability, it would be to the advantage of the government to accept the VECP provided that the costs of implementing the change are not large.

Pricing Engineering Change Proposals

The ACO sometimes negotiates agreements incorporating approved engineering change proposals. The initial step is to measure the extent of the change. For this purpose, the best measurement may be the percentage of changed effort. Either work standards, if they are available, or engineering estimates may be used. Such estimates are based on experience with similar components. The nature of the estimating systems used varies widely from company to company. One simple and crude estimate of the percentage of change is to relate the number of blueprints to be changed with

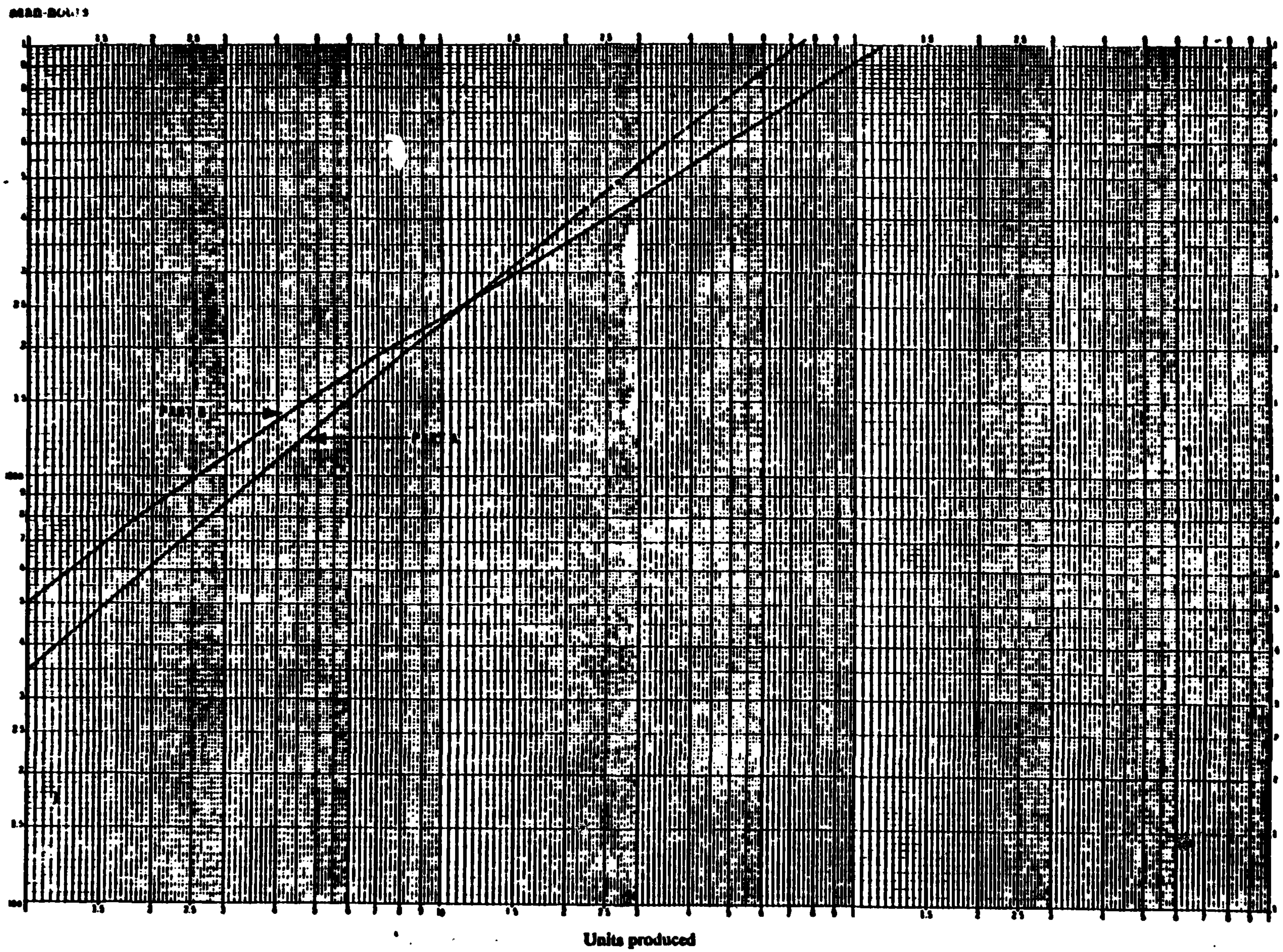


Figure 3-7 The learning curve and cost volume relationship.

the total number of blueprints covering the entire item or component. To use the learning curve appropriately, the estimator must be able to express the engineering change proposal by a percentage in terms of the total work to be done.

A phenomenon of the learning curve is this: when change is introduced, learning tends to begin anew. In other words, if all production required new learning, the man hour reduction following the first unit and the resulting efficiencies would be lost. Under this concept, the amount of labor required for changing the first unit agrees in percentage with the amount of change.

The following example assumes an instance in which one equal part is being substituted for another during the production of an end item. The change affects the units between number 300 and number 500, a total of 200 units. The change deletes 10 percent of the old work and adds 10 percent new work. The procedure, using learning curve tables, for computing the value in hours of the changed portion of the work is as follows:

First, consult the learning curve cumulative total tables for the rate of 500 units and the rate of 300 units on the 80 percent learning curve. (Figure 3-8).

500 units = 98.847
300 units = 69.663
200 units = 29.184

Second, apply the formula F times R to find the total number of man hours which would be expended if no change occurred.

F equals the man hour value of 1st unit = 500 man hours
R equals the rate for R_{500} minus R_{300} = 29.184

500 x 29.184 = 14,592 manhours for 200 units

Third, apply the same formula to the amount of unchanged work.
90 percent X F = 450 man hours
450 man hours X 29.184 = 13,133 man hours for 200 units

Fourth, calculate the value in hours for the changed portion of the work. F for the changed portion of the work equals 10 percent times 500 or 50 man hours. Assume the same slope for learning on the changed portion of the work as for the unchanged portion of the work.

R for 200 units of the changed portion starting at unit 1 = 52.720
50 man hours X 52.720 = 2,636 man hours

Fifth, add the amount of unchanged work to the amount of changed work for the total number of man hours for 200 units.

13,133 + 2,636 = 15,769 man hours

Figure 3-8
Segment of a Learning Curve Rate Table

80% Curve

Unit Number	Cumulative Total	Cumulative Average	Unit
1	1.00000000	1.00000000	1.00000000
10	6.31537300	0.63153730	0.47650987
20	10.48494279	0.52424714	0.38120790
30	14.01989341	0.46732978	0.33455934
40	17.19345570	0.42983639	0.30496632
50	20.12171364	0.40243427	0.28382707
300	69.66337918	0.23221126	0.15942083
500	98.84724744	0.19769449	0.13524640

Extract from: The Experience Curve Tables, Vols. 1 and 2, U. S. Army Missile Command, Redstone Arsenal, Alabama, 1962.

Sixth, the cost of the change is the difference in the total man hours of the changed procurement and the total man hours of the original procurement.

15,769 - 14,592 = 1,177 added man hours (cost of the change)

This example concerned the substitution of one equal part for another in the end item. Other changes may cause either an addition to or a deletion from the end item. If work is added, the procedure is the same except that no work is deleted. If the change resulted in a deletion, the percentage of old work would be reduced, and no new work would be added.

CHAPTER 4

MAINTENANCE

Maintenance is the process employed to keep equipment in condition for effective use or to return it to that condition when it malfunctions or fails. It involves repair, overhaul, servicing, calibration, inspection, modification, and manufacture. The Air Force delineates five primary objectives of maintenance (AFR 66-14): (1) Develop and maintain the military and industrial capability needed to sustain Air Force operations under all conditions; (2) Insure organizational structures are designed on the basis of a wartime operational concept to provide an in-being maintenance support capability to meet all operational requirements; (3) Establish Air Force maintenance systems and techniques that will be responsive to changing operational requirements and technology; (4) Insure that Air Force material is serviceable, operable, and configured to meet the mission requirements; and (5) Insure that maintenance planning begins in the conceptual phase of each new system or equipment. The airlines define their objectives for maintenance more succinctly, and from a different perspective, viz., "To prevent deterioration of the inherent design levels of reliability and operating safety of the aircraft and to accomplish this protection at the minimum practical costs."

In the Air Force, equipment maintenance is the largest facet of logistics in terms of money, manpower, facilities, and many other important resources. As such, maintenance is considered a major factor in military capability.

The maintenance system may be considered as a "production system" and is generally operated in parallel with the organization's primary production system. Maintenance management, therefore, employs essentially the same management tools and techniques as found in the production/operations management area. Because of the complexity of many of the problems found in maintenance, heuristics (e.g., "Johnson's Rule" for scheduling) are used extensively. Simulation is another tool which is finding wider application in the maintenance area. The problems illustrated in this section employ techniques which can be found in any comprehensive, senior undergraduate or beginning graduate level text on production/operations management.

4-1 ASSEMBLY LINE BALANCING

An emergency TCTO on the F-15 requires seventeen tasks to complete. The tasks, designated by letter, their performance times, and sequence, are listed in Table 4-1. A fleet of 72 aircraft must be run through the TCTO process. The DCM wants the job completed in no more than six days. The wing works two shifts per day, seven productive hours per shift.

The DCM wants to set up an assembly line type of operation. How many work stations should he use? What tasks do they consist of?

Table 4-1

TASKS, PERFORMANCE TIMES, AND PRECEDENCE

TASK	PERFORMANCE TIME (MIN)	TASKS THAT MUST PRECEDE
A	10	--
B	5	A
C	15	A,B
D	15	--
E	30	D
F	15	E,F
G	45	A,B,C,D,E,F
H	15	A,B,C
I	20	D,E,F
J	25	A,B,C,D,E,F,G,H
K	10	D,E,F,I
L	25	D,E,F,I
M	40	A,B,C,D,E,F,G,H,J
N	25	D,E,F,I,K
O	30	D,E,F,I,L
P	50	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,
Q	15	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P
	390	

This is an assembly line balancing problem. The layout of the assembly line, in terms of the number of work stations required, is based on the output required from the line and the desire for an efficient operation, i.e., one that gets the job done, on time, with a minimum amount of idle or wasted time.

The first problem, then, is to determine the output required, or demand (D). Since there are 72 aircraft, and they must all be completed in no more than six days, demand is:

$$D = \frac{72 \text{ acft}}{6 \text{ days}} = 12 \text{ acft/day}$$

Next, we need to know the productive time (P) per day available to perform the TCTO. This time is:

$$P = 2 \text{ shifts/day} \times 7 \text{ hours/shift} = 14 \text{ hrs/day}$$

From Table 4-1, the total time (T) required for the seventeen tasks in the TCTO is 390 minutes, or 6.5 hours.

The number of work stations required (N) is then:

$$N = \frac{TXD}{P} = \frac{6.5 \frac{hrs}{acft} \times 12 \frac{acft}{day}}{14 \frac{hrs}{day}} = 5.57$$

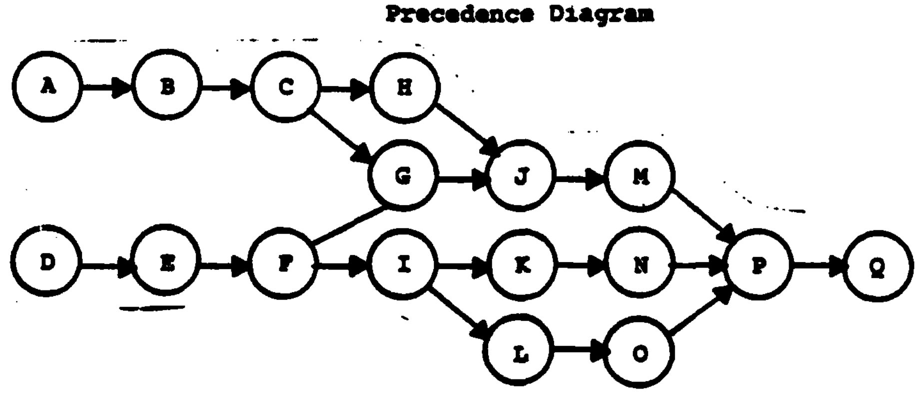
Therefore: six work stations are required.

The cycle time (C) required to meet the demand, i.e., the maximum time permissible between aircraft coming off the assembly line, is P/D = 14/12 = 1.167 hrs, or 70 minutes. Since the maximum task time, task P, is 50 minutes, we can meet the demand with a properly balanced line.

In order to see what this line would look like, we need to allocate tasks to different work stations, within the given precedence constraints. In order to better visualize these precedence constraints, a precedence diagram is drawn from the data in Table 4-1. This diagram is shown in Figure 4-1.

Figure 4-1

Precedence Diagram



To make the first line balance, we will use a rule which states: "First allocate those tasks that have the longest operation time." In case of a tie, we will allocate the task with the largest number of following tasks. The results of employing this rule are shown in Table 4-2.

Table 4-2

Balance Made According to Longest Operating Time (70 Min Cycle Time)

TASK	TIME	REMAINING UNASSIGNED TIME (MIN)	FEASIBLE REMAINING TASKS
D	15	55	A,E
E	30	25	A,F
F	15	10	A
A	10	0*	None
I	20	50	B,K,L
L	25	25	B,K
K	10	15	B
B	5	10*	None
O	30	40	N,C
N	25	15	C
C	15	0*	None
G	45	25	H
H	15	10*	None
J	25	45	M
M	40	5*	None
P	50	20	Q
Q	15	5*	None
		30**	

*idle time per station
**total idle time for line

From the table, we see that indeed six work stations can be structured and result in a line with 30 minutes total idle time (Figure 4-2). The efficiency of this line may be computed as:

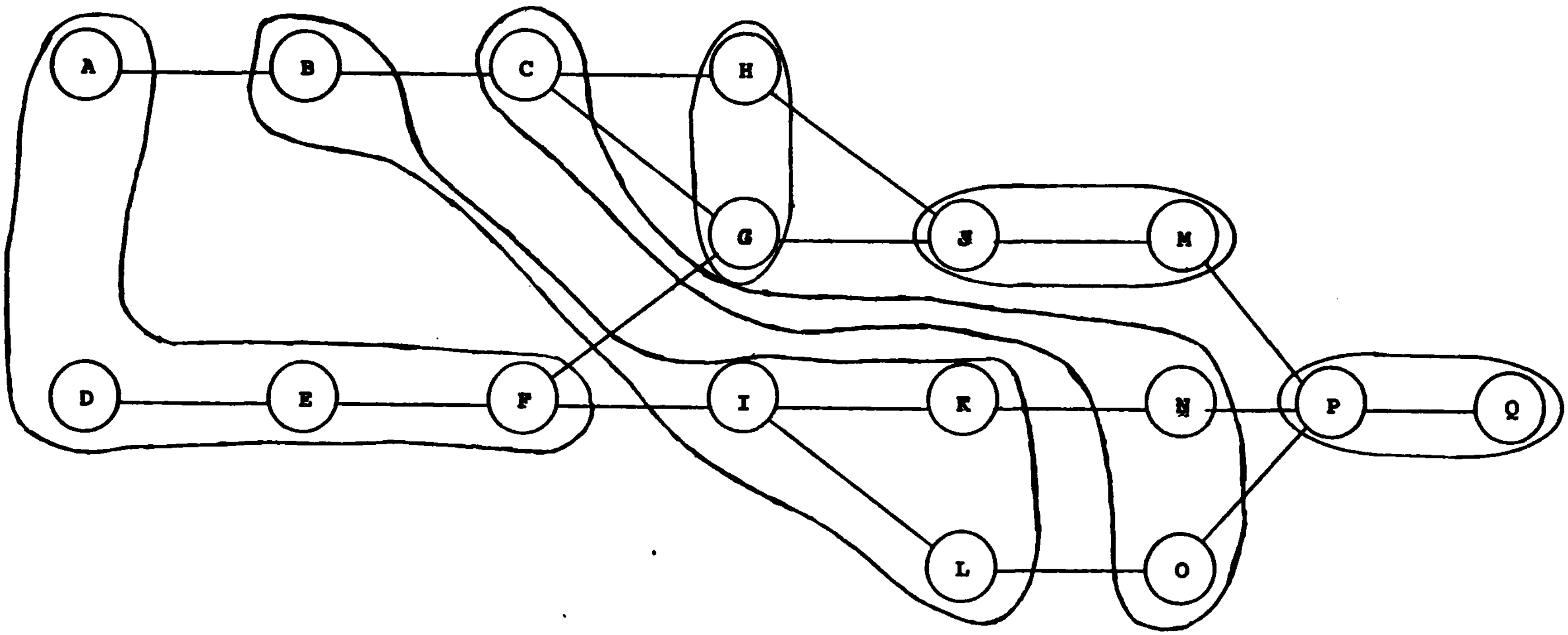
$$\text{Efficiency (\%)} = \frac{390}{390 + 30} \times 100 = 93\%$$

also,

$$\text{Efficiency} = \frac{\sum_{t=1}^{17} t_i}{NC} = \frac{390}{6(70)} = .93$$

Figure 4-2

Work Stations for Balance shown in Table 4-2



4-2 SCHEDULING JOBS THROUGH ONE PROCESS OR MACHINE

The supervisor of the "wash rack" facility (WRF) at Cormorant Air Logistics Center has been receiving complaints from the Component Repair Section (CRS) that jobs are taking too long on the average to go through the steam cleaning process. This delays component overhaul and return to the customer. The CRS supervisor claims the mean flow time through the steam cleaning process has been over 18 hours and he must have a mean flow time of less than 12 hours if he is to increase the CRS's probability of meeting its schedule.

The WRF supervisor examined his records and found that the same 10 jobs accounted for 80 to 90 percent of his CRS work load. These jobs and their typical processing time for steam cleaning are shown in Table 4-3, below.

Table 4-3

Job	Steam Cleaning Time (Hrs)
ACQ 121	2.25
BRC 154	1.75
CRC 5S	0.75
DNA 111	0.50
ELEC 6A	3.75
FQR 3	1.25
GIM JT	2.50
HYD 1A	6.00
IND 7	3.00
JP 4A	3.25

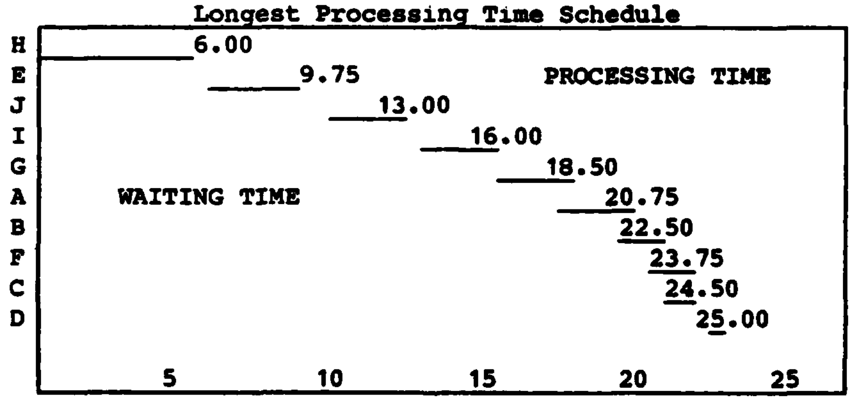
The normal procedure was for the WRF supervisor to get all the job orders at the beginning of the week for that week's work. He would then schedule the jobs that his experience told him were most difficult and time-consuming first, in order to "get them out of the way." He would then schedule the remaining easier jobs. Now he was being told that the jobs were, on the average, taking too long. He wondered how he might minimize the mean flow time through his steam cleaning process.

This is a job shop scheduling situation, with a static job arrival pattern. It falls into that class of problem termed an "n job--one-machine problem." In this type of problem, the shortest processing time rule <u>always</u> yields the optimum solution.

In this case, the WRF supervisor is actually scheduling the jobs with the longest processing time first. Thus, he would schedule the jobs in Table 4-3 according to the sequence shown in Figure 4-3. The mean flow time for this schedule is computed by summing the flow times for all jobs; i.e., 6.00 + 9.75 + 13.00 + 16.00 + 18.50 + 20.75 + 22.50 + 23.75 + 24.50 + 25.00, and dividing by the number of jobs, 10, i.e., 179.75/10 = 17.975. Thus, the mean flow time for this scheduling sequence is almost 18 hours, well over the 12-hour mean requested by the CRS supervisor. Using the shortest processing time rule, the jobs would be sequenced as shown in Figure 4-4 and the minimun mean flow time would be: (.50 + 1.25 + 2.50 + 4.25 + 6.50 + 9.00 + 12.00 + 15.25 + 19.00 + 25.00) ÷ 10 = 9.525. Thus, use of the shortest processing time rule almost halves mean flow time and allows the WRF supervisor to meet the requirement for less than 12 hours mean flow time.

Figure 4-3

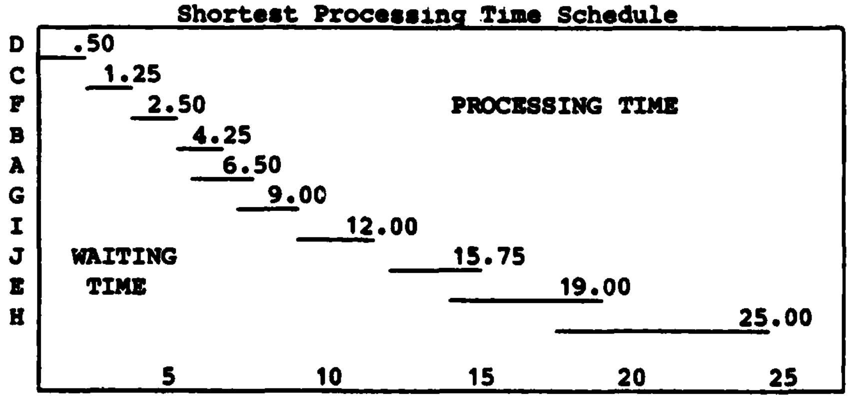
Longest Processing Time Schedule



TIME (HRS)

Figure 4-4

Shortest Processing Time Schedule



TIME (HRS)

4-3 SCHEDULING n JOBS THROUGH m MACHINES

The supervisor of the Component Repair Squadron (CRS) at Korona AB Centralized Intermediate Repair Facility (CIRF) was reviewing the jobs he had to schedule for the upcoming week. He was concerned because operating bases were complaining about the slow turnaround they were getting on their repair requests. The CRS supervisor had been told to get the current batch of jobs out ASAP and not later than within 48 hours. The CRS supervisor wondered whether by proper scheduling he might get all the jobs through the shop within the time required, using only the men and equipment available. The jobs and the processing time for each job on each of three machines they must be run through are listed in Table 4-4, below. Each job must go across machine I first, then machine II, then III, in that order.

Table 4-4

JOB	MACHINE (HRS) I	II	III	TOTAL
AQN	6	1	9	16
BRD	8	5	7	20
CRC	10	5	6	21
DNI	7	6	11	24
EOQ	5	3	6	14
				95
	Min=6	Max=6	Min=6	

This is an n/3 flow shop scheduling situation, with a static job arrival pattern. In this type of problem, an optimal solution can be obtained by an extension of "Johnson's Rule" for the n/2 case.

In this problem, sum job processing times on Machines I and II, then sum job processing times on Machines II and III. This will give two columns of times as shown in Table 4-5 below. Apply "Johnson's Rule" to this tableau, to determine job sequence*.

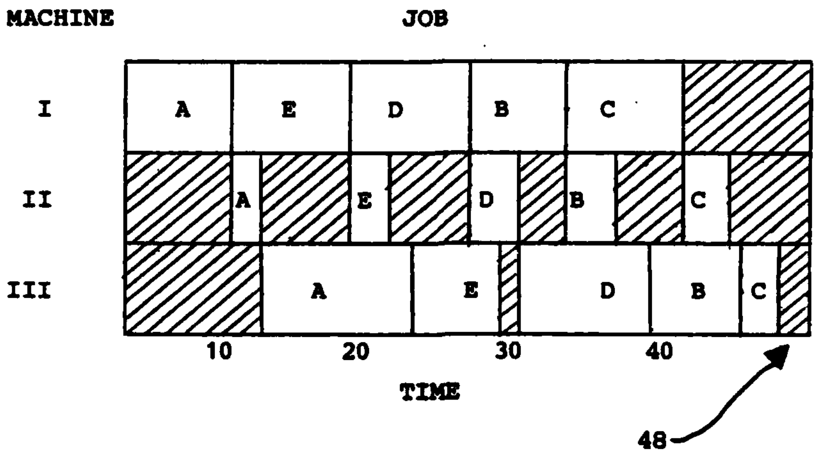
Table 4-5

JOB	MACHINE (HRS) I & II	II & III	SEQUENCE BY JOHNSON'S RULE
AQN	7	10	1
BRD	13	12	4
CRC	15	11	5
DNI	13	17	3
EOQ	8	9	2

Plotting the jobs in the predetermined sequence across each machine as shown in Figure 4-5, below, indicates that the jobs may be completed in 48 hours.

Figure 4-5

Sequence of Jobs on Machines



*It should be noted that Johnson's Rule will give an optimum solution for the n-job, 3-machine problem only where the maximum operation time on the second machine is less than or equal to the minimum operation time on either of the other machines, i.e.,

$$\min t_{i1} \geq \max t_{i2} \qquad \text{or} \qquad \min t_{i3} \geq \max t_{i2}$$

Where t_{ij} = processing time of the ith job on the jth machine.

4-4 USE OF THE LEARNING CURVE IN ESTIMATING MAINTENANCE MANPOWER REQUIREMENTS

A contractor, selected to produce a new aircraft for the USAF, has indicated to the government that the aircraft can be expected to require 20 maintenance man-hours per flying hour (MH/FH) approximately 18 months after it is introduced into the operational Air Force. During Operational Test and Evaluation, the first aircraft off the line experienced approximately 100 MH/FH's. You've been asked to evaluate the contractor's MH/FH projection, subsequent to that MH/FH factor being used to estimate downstream maintenance manpower requirements. The contractor has indicated that the aircraft is similar to the F-222, currently in the USAF active inventory. Data for the first 18 months of operation for the F-222 is shown in Table 4-6 below.

Table 4-6

Flying Hours Per Month	Man-Hours Per Month
108	9,601
187	19,298
256	17,894
490	28,322
800	42,480
1106	55,189
1858	91,042
1927	86,715
1801	79,064
2700	119,880
3605	144,560
3659	145,262
4800	168,480
6400	204,800
7208	223,448
8000	272,000
7982	278,572
8640	285,120

There are a number of ways to attack this problem. Data is scanty, and numerous assumptions must be made. Whatever figure is finally derived for a MH/FH factor will have to be considered a preliminary estimate, subject to further refinement as more data becomes available.

We'll take a learning curve approach to this problem. Let's examine the contractor's claims in terms of what the learning factor has to be to achieve a 20 MH/FH factor in the 18th month of operation, given the 1st month's 100 MH/FH factor. The expression for the Learning Curve is

$$Y_X = KX^n$$

Where:

Y_X = man-hours/flying hour for the X^{th} unit,

K = " " " " " 1st " ,

X = the X^{th} unit

n = log b/log 2

b = learning factor

For our case:

$$20 = 100(18)^{\log b/\log 2}$$

$$\log .20 = \log b/\log 2 \,(\log 18)$$

$$\log b = -.1676$$

$$b = .68$$

Thus, the contractor is expecting a 68% learning factor--rather optimistic relative to what the aircraft industry experiences, which is closer to 80%.

Looking at the data for the F-222 (Table 4-6), one can see that as the number of flying hours per month increases, the number of man-hours per month increases as well, but at a decreasing rate. We can analyze this phenomenon using learning curve theory by reducing the man-hours per month to man-hours per flying hour and comparing the month with the MH/FH figure for that month. We can use regression analysis to derive a learning factor for this transformed data. The month represents the independent variable, and the MH/FH's the dependent variable. the expression we'll be looking for will be the log transformation of the basic learning curve equation, i.e.,

$$Y = KX^n$$
$$\log Y = \log K + n \log X$$

which takes the form

$$Y = a + b\,X$$

The data pairs used in deriving this linear bivariate model are thus the logs of the X and Y variables shown in Table 4-7, below.

Table 4-7

X & Y Variables

X	Y
1	88.9
2	103.2
3	69.9
4	57.8
5	53.1
6	49.9
7	49.0
8	45.0
9	43.9
10	44.4
11	40.1
12	39.7
13	35.1
14	32.0
15	31.0
16	34.0
17	34.9
18	33.0

Using a standard Linear Regression program, such as found in Statistical Package for the Social Sciences results in the expression:

$$Y = 2.0293 - .4154X$$

Remember this as: $\log Y = \log K + n \log X$

We translate to: $\log K = 2\ 0293$

and, $n = \log b/\log 2 = -.4154$

then, $\log b = -.1250$

and, $b = .75$

If the contractor's new aircraft is similar to the F-222, perhaps a Learning Factor of 75% would be a more reasonable estimate than the 68% the contractor is implying. If we use 75% for the new aircraft, we

find for the 18th month:

$$Y_{18} = 100(18)^{\log .75/\log 2}$$

$$Y_{18} = 100(.3013) = 30.13$$

Thus, for initial planning, a MH/FH factor of approximately 30 for the 18th month of operation appears to be a better estimate than the 20 MH/FH's the contractor has predicted.

What impact does a 10 MH/FH change have on aircraft maintenance manpower estimates?

Aircraft maintenance manpower requirements can be computed based on 4 factors:

(1) The MH/FH factor for the weapon system under study, e.g., the F-222.

(2) The Unit Equipment (UE)--number of aircraft possessed--for the squadron, wing, etc., under consideration.

(3) the programmed flying hours (FH) per aircraft.

(4) The productive direct manhours available per month per man (AM).

Manpower (M) may be computed as follows:

$$M = \frac{MH/FH \times FH \times UE}{AM}$$

For example, suppose the F-222 has a MH/FH of 30, flies 25 hours/month and is assigned to 6 wings in the USAF with a UE of 72 aircraft per wing. Productive direct available manhours per man per month are 86.4. (This is a standard figure based on 60% of 144 available manhours per month.) Then,

$$M = \frac{30 \text{ MH/FH X } 25 \text{ FH/MO X } 72 \text{ ACFT}}{86.4 \text{ MH/MO}} = 625$$

For 6 wings,

$$M_6 = 6 \text{ X } 625 = 3750$$

Generally, certain factors are applied to this figure to account for functions not covered by the MH/FH factor, such as, the Chief of Maintenance Function, Aerospace Ground Equipment, Survival Equipment, etc. The aggregate of these factors can amount to 117% of the productive

direct manpower figure, or in our example,

$$M_6 = 3750 \text{ X } 1.17 = 4388$$

At 20 MH/FH, instead of 30 MH/FH, our figure would be,

$$M_6 = \frac{20 \text{ X } 25 \text{ X } 72}{86.4} \text{ X } 6 \text{ X } 1.17 = 2925$$

a difference of,

$$4388 - 2925 = 1463$$

manpower authorizations--over 240 per wing! This could make a big difference in wing maintenance capability. Perhaps more importantly at this stage of manning, it makes a substantial difference in budget planning. At about $12,000 per authorization, a difference of 463 authorizations equates to about $17.6 million.

4-5 BASE LEVEL EMERGENCY INSPECTION OF THE T-37 FLEET

On 26 February 1979, Air Training Command (ATC) sent an emergency message to all Undergraduate Pilot Training (UPT) bases (see Exhibit 1). The message restricted all T-37 aircraft from flight in excess of +3.0 Gs and less that -1.5 Gs. It also prohibited aerobatic maneuvers exceeding these G tolerances and spins. This action resulted from a discovery by the San Antonio Air Logistics Center (SAALC) that a T-37 undergoing normal programmed depot maintenance had two cracked wing spar bolts and one cracked spar attachment fitting (see Exhibit 2). The seriousness of the situation was self-evident; any T-37 with similar structural discrepancies risked wing separation during high or negative G maneuvering. The required corrective action was a one-time inspection of all T-37 aircraft, to be accomplished at base level. If base maintenance found the wing spar bolts or attachment fittings cracked, they were to be replaced at the base. The SAALC did not have an estimate of how many T-37s actually had wing spar bolts or attachment fittings cracked. ATC tasked Reese AFB with constructing a plan to be used by all UPT bases for efficient accomplishment of the one-time inspection and corrective action, if necessary. The plan was to be submitted to ATC by 10 March, and the inspection was to be started on 15 March.

The activities required for the inspection are shown in Exhibit 3. Placing the wings in a no load position (activity 4) requires a set of wing cradles. Each ATC T-37 base was allocated only one set of such cradles.

A survey of the bases involved in the inspection was accomplished to determine the number of spare bolts and fittings, aircraft possessed, and personnel manning. This information is shown in Exhibit 4. How should this inspection be scheduled? How long will it take to complete?

This problem involves the use of activity charting, following some simple computations to determine number of inspection docks to employ. The constraining factor in the inspection would appear to be the wing cradles. Since each base has only one set of cradles, it is necessary to minimize the amount of time they are on any one aircraft. Referring to Exhibit 3, the sequence of activities involving the cradles appears to be as shown below.

	Min	
Place wings in no load	30	
Remove bolts and fittings	15	
Inspect bolts and fittings	30	
Replace bolts and fittings	15	150 Min.
QC Inspection	30	
Remove wings from no load	30	

EXHIBIT 1: THE EMERGENCY MESSAGE

SUBJECT: T-37 WING SPAR BOLT, NUT, AND ATTACHMENT FITTINGS.

(1.) T-37 TAIL NUMBER 57-2344 WHILE UNDERGOING PDM AT SAALC WAS FOUND TO HAVE TWO WING SPAR BOLTS AND ONE WING SPAR ATTACHMENT FITTING WHICH WERE CRACKED AND REQUIRED REPLACING. DUE TO THIS OCCURRENCE, A ONE-TIME INSPECTION OF ALL T-37 WING SPAR BOLTS, NUTS, AND ATTACHMENT FITTINGS WILL BE ACCOMPLISHED NLT 180 DAYS FROM THE RECEIPT OF THE MSG. UNTIL THIS INSPECTON IS ACCOMPLISHED, ALL T-37s ARE RESTRICTED TO +3.0 Gs and -1.5 Gs, AND AEROBATICS EXCEEDING THESE TOLERANCES AND SPINS ARE PROHIBITED.

(2.) WING SPAR BOLTS AND ATTACHMENT FITTINGS WILL BE INSPECTED IAW IT-378-6, PARA 6-14. CRACKED BOLTS WILL BE REPLACED IAW IT-37B-2-2, PARA 3-16. CRACKED ATTACHMENT FITTINGS WILL BE REPLACED IAW IT-37B-2-6, PARA 3-12.

(3.) PRESENTLY SAALC HAS 136 WING SPAR BOLTS (PN 500675-4) AND 48 ATTACHMENT FITTINGS (PN 500855-1) IN SUPPLY. REQUIREMENTS EXCEEDING BASE STOCKS SHOULD BE ORDERED ASAP.

(4.) REESE AFB, TX WILL DESIGN THE INSPECTION PLAN TO BE USED BY ALL BASES. THE PLAN MUST BE FURNISHED TO HQ ATC/LG NLT 10 MARCH 79. THE PLAN WILL BE DISSEMINATED TO ALL BASES ON 12 MARCH AND INSP WILL BEGIN ON 15 MARCH.

(5.) REESE AFB, TX - AFTER INSPECTION OF 10 AIRCRAFT, YOU WILL PROVIDE ATC/LGS WITH A FORECAST OF THE EXPECTD WING SPAR BOLTS AND NUTS REQUIREMENT FOR YOUR BASE. THIS FORECAST WILL BE USED TO REQUEST ADDITIONAL SPAR NUTS AND BOLTS THROUGH SUPPLY CHANNELS.

(6.) THE ATC OPERATIONAL READY (OR) RATE OF 70% MUST BE MAINTAINED.

(7.) REESE AFB IS AUTHORIZED TO START INSPECTION IMMEDIATELY IN ORDER TO AID IN THE DEVELOPMENT OF THE MODEL PLAN. FINDINGS AND SOLUTIONS SHOULD BE INCLUDED IN THE MODEL PLAN.

EXHIBIT 2

VISUAL OF REQUIRED INSPECTION
AREA (WINGS IDENTICAL)

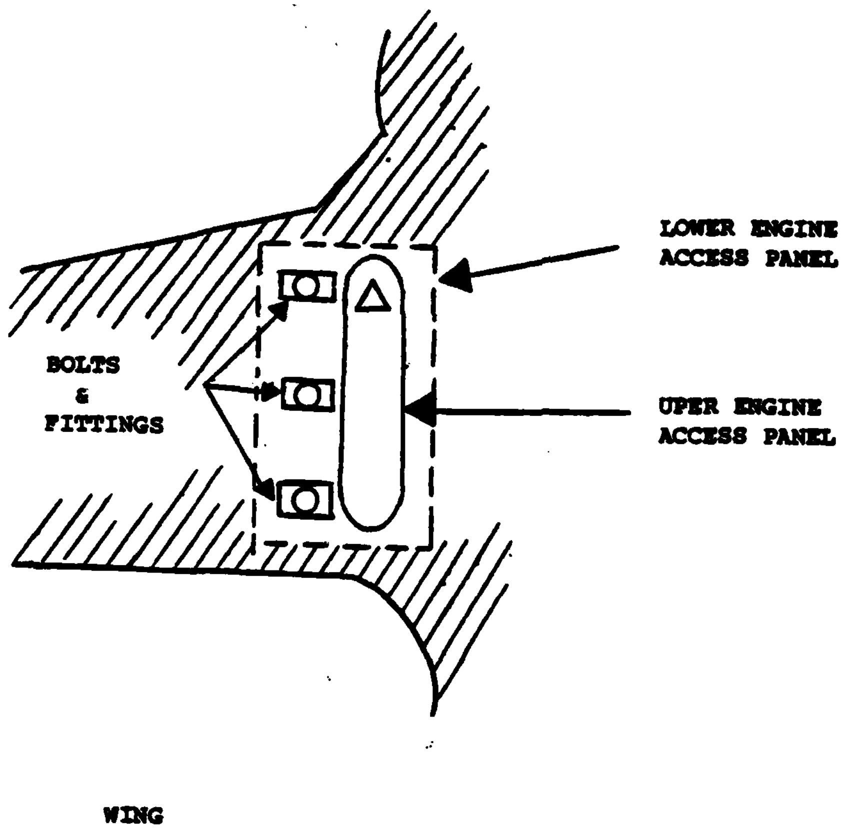


EXHIBIT 3

T.O. IT-378-6, PARA 6-14. Inspection of Wing Spar Bolts and Attachment Fittings.

Activities	Time Required (Min)	Personnel Required Skill	Personnel Required Quantity
1. Jack Aircraft	15	APG	2
2. Remove Engine Panels	15	APG	2
3. Remove Engines	120	ENG HYD	2 1
4. Place Wings in No-Load	30	MECH	2
5. Remove Bolts & Fittings	15	MECH	2
6. Inspect Bolts & Fittings	30	NDI	1
7. Replace Bolts & Fittings	15	MECH	2
8. QC Inspection	30	QC	1
9. Remove Wings from No-Load	30	MECH	2
10. Install Engines	120	ENG HYD	2 1
11. Replace Panels	15	APG	2
12. Down Jack Aircraft	15	APG	2

Notes:

1. Inspection will take place in hangar.

2. Towing into and out of hangar takes 15 min., and requires 2 APG personnel, and 2 other personnel of any skill type.

3. Aircraft will require engine operational check and FCF prior to being returned to FTS. This takes one full day, on the average.

EXHIBIT 4

Base	Aircraft Possessed	Spare		Personnel Assigned					
		Bolts	Fittings	APG	ENG	HYD	MECH	NDI	QC
Columbus	48	8	3	54	9	5	6	3	3
Laughlin	36	6	2	43	7	3	5	2	2
Reese	72	12	4	82	10	7	7	3	3
Vance	48	8	4	51	10	3	6	3	3
Williams	72	10	3	88	11	6	5	3	3
Sheppard	48	6	2	56	9	5	5	3	3

Notes: ATC bases work three, 8-hour shifts, per day (24 hours per day, 5 days per week.

The wing cradles are tied up for 150 minutes. Therefore, the maximum number of aircraft we can process through inspection each day is:

$$\frac{(24\ \text{Hr/Day})\ (60\ \text{Min/Hr})}{150\ \text{Min/Acft}} = 9.6$$

However, 30 minutes of the time the cradles are employed is, in effect, idle time, since the bolts and fittings are inspected while the aircraft is in jig. If the bolts and fittings are immediately replaced with good spares, the time required for cradle employment can be reduced 20% (from 150 to 120) and the maximum number of aircraft that can be processed through inspection each day is:

$$\frac{(24\ \text{Hr/Day})\ (60\ \text{Min/Hr})}{120\ \text{Min/Acft}} = 12.0$$

However, the total time required for the inspection (not including the engine ops check or aircraft FCF) is 450 minutes. Therefore, the maximum number of aircraft that can be processed through inspection in any 24-hour period is:

$$\frac{24 \times 60}{450} = 3.2$$

Therefore, in order to complete the inspection on 12 aircraft per day, we require 12/3 = 4 inspection docks. In order to determine the sequence in which the aircraft should be worked, and assess the personnel impacts, an activity chart can be constructed, as shown in Table 4-8.

From an examination of the activity chart, it can be seen that the number of personnel required per shift and the total personnel requirements by skill, are as shown in Table 4-9, below.

Table 4-9

Personnel Requirements by Skill

SKILL	PERSONNEL/SHIFT	TOTAL PERSONNEL FOR 3 SHIFTS
APG	2	6
ENG	4	12
HYD	2	6
MECH	2	6
QC	1	3
NDI	1	3

Table 4-8

ACTIVITY CHART

T	DOCK 1 ACTIVITY	DOCK 2 ACTIVITY	DOCK 3 ACTIVITY	DOCK 4 ACTIVITY	AEHMQN PNYECD GGDC I
15	TOW		REM WINGS		242 1
30	JACK		FM NO LOAD		2422
45	DEPANEL				2422
	REMOVE	REPL PANELS	INSTALL	NO LOAD WINGS	2422
	ENGINES	DOWN JACK	ENGINES		2422
		TOW		REM BLTS/FIT	2422
				REPL BLTS/FIT	422 1
				QC INSP	42 11
		TOW			242 1
		JACK		REM WINGS	2422
165		DEPANEL		FM NO LOAD	2422
	NO LOAD WINGS	REMOVE	REPL PANELS	INSTALL	2422
195		ENGINES	DOWN JACK	ENGINES	2422
210	REM BLTS/FIT		TOW		2422
225	REPL BLTS/FIT				422 1
	QC INSP				42 11
255			TOW		242 1
	REM WINGS		JACK		2422
285	FM NO LOAD		DEPANEL		2422
	INSTALL	NO LOAD WINGS	REMOVE	REPL PANELS	2422
	ENGINES		ENGINES	DOWN JACK	2422
		REM BLTS/FIT		TOW	2422
		REPL BLTS/FIT			422 1
		QC INSP			42 11
				TOW	242 1
		REM WINGS		JACK	2422
405		FM NO LOAD		DEPANEL	2422
420	REPL PANELS	INSTALL	NO LOAD WINGS	REMOVE	2422
435	DOWN JACK	ENGINES		ENGINES	2422
450	TOW		REM BLTS/FIT		2422
			REPL BLTS/FIT		422 1
480			QC INSP		42 11
495	TOW				242 1
510	JACK		REM WINGS		2422
525	DEPANEL		FM NO LOAD		2422
	REMOVE	REPL PANELS	INSTALL	NO LOAD WINGS	2422
	ENGINES	DOWN JACK	ENGINES		2422
		TOW		REM BLTS/FIT	2422
				REPL BLTS/FIT	422 1
				QC INSP	42 11
		TOW			242 1
		JACK		REM WINGS	2422
645		DEPANEL		FM NO LOAD	2422
	NO LOAD WINGS	REMOVE	REPL PANELS	INSTALL	2422
675		ENGINES	DOWN JACK	ENGINES	2422
690	REM BLTS/FIT		TOW		2422
705	REPL BLTS/FIT				422 1
	QC INSP				42 1

Table 4-8 (cont.)

735			TOW		242 1
	REM WINGS		JACK		2422
765	FM NO LOAD		DEPANEL		2422
	INSTALL	NO LOAD WINGS	REMOVE	REPL PANELS	2422
	ENGINES		ENGINES	DOWN JACK	2422
		REM BLTS/FIT		TOW	2422
		REPL BLTS/FIT			422 1
		QC INSP			42 11
				TOW	242 1
		REM WINGS		JACK	2422
885		FM NO LOAD		DEPANEL	2422
900	REPL PANELS	INSTALL	NO LOAD WINGS	REMOVE	2422
915	DOWN JACK	ENGINES		ENGINES	2422
930	TOW		REM BLTS/FIT		2422
			REPL BLTS/FIT		422 1
960			QC INSP		42 11
975	TOW				242 1
990	JACK		REM WINGS		2422
1005	DEPANEL		FM NO LOAD		2422
	REMOVE	REPL PANELS	INSTALL	NO LOAD WINGS	2422
	ENGINES	DOWN JACK	ENGINES		2422
		TOW		REM BLTS/FIT	2422
				REPL BLTS/FIT	422 1
				QC INSP	42 11
		TOW			242 1
		JACK		REM WINGS	2422
1125		DEPANEL		FM NO LOAD	2422
	NO LOAD WINGS	REMOVE	REPL PANELS	INSTALL	2422
1155		ENGINES	DOWN JACK	ENGINES	2422
1170	REM BLTS/FIT		TOW		2422
1185	REPL BLTS/FIT				422 1
	QC INSP				42 11
1215			TOW		242 1
	REM WINGS		JACK		2422
1245	FM NO LOAD		DEPANEL		2422
	INSTALL	NO LOAD WINGS	REMOVE	REPL PANELS	2422
	ENGINES		ENGINES	DOWN JACK	2422
		REM BLTS/FIT		TOW	2422
		REPL BLTS/FIT			422 1
		QC INSP			42 11
				TOW	242 1
		REM WINGS		JACK	2422
1365		FM NO LOAD		DEPANEL	2422
1380	REPL PANELS	INSTALL	NO LOAD WINGS	REMOVE	2422
1395	DOWN JACK	ENGINES		ENGINES	2422
1410	TOW		REM BLTS/FIT		2422
			REPL BLTS/FIT		422 1
1440			QC INSP		42 11

END OF DAY

Comparing this to personnel assigned by base, as shown in Exhibit 4, gives us the figures in Table 4-10. From an examination of this table, we can see that there are several skill shortages among the bases. Thus, we cannot operate a 4-dock, 3-shift per day operation at every base.

In order to determine what our options are, we can look at personnel requirements for different operating set-ups. This is done in Table 4-11 where personnel requirements by skill are shown for different combinations of docks and shifts per day. As can be seen from this table, there is no advantage to running only 3 docks, since personnel requirements for all skills are the same. Requirements are reduced in Engine and Hydraulic skills when only two docks are operated. Reducing the number of shifts reduces requirements in every case.

Next, we look at the number of days required to complete the inspection for different combinations of docks and shifts per day, for different numbers of aircraft possessed, i.e., 36, 48, or 72. Examination of Table 4-12 indicates that the minimum amount of time required to process 72 aircraft through the inspection is 6 days. If we decide to complete the inspection for all bases in 6 days, we would operate 4 docks, 3 shifts per day at Reese and Williams, which possess 72 aircraft; 4 docks, 2 shifts per day at Columbus, Vance, and Sheppard, which possess 48 aircraft; and 2 docks, 3 shifts per day at Laughlin, which possesses 36 aircraft. In this way, the inspection would be completed in 6 days at all ATC bases.

Reassessing the personnel impacts results in the shortages and overages shown in Table 4-13. In order to complete the inspection in 6 days at all bases, some shifts in personnel assets will be required. For example, both Reese and Williams are short of engine specialists, however, all other bases have excess specialists. Therefore, the option of sending excess specialists TDY to bases with shortages should be explored.

Table 4-10

REQUIRED VERSUS ASSIGNED BY SKILL FOR 4 DOCKS, 3 SHIFTS PER DAY

SKILL \ BASE		COL (48)	LAU (36)	REE (72)	VAN (48)	WIL (72)	SHE (48)
APG	RQD	6	6	6	6	6	6
	ASGN	54	43	82	51	88	56
	+/-	+48	+37	+76	+45	+82	+50
ENG	RQD	12	12	12	12	12	12
	ASGN	9	7	10	10	11	9
	+/-	- 3	- 5	- 2	- 2	- 1	- 3
HYD	RQD	6	6	6	6	6	6
	ASGN	5	3	7	3	6	5
	+/-	- 1	- 3	+ 1	- 3		- 1
MECH	RQD	6	6	6	6	6	6
	ASGN	6	5	7	6	5	5
	+/-		- 1	+ 1		- 1	- 1
QC	RQD	3	3	3	3	3	3
	ASGN	3	2	3	3	3	3
	+/-		- 1				
NDI	RQD	3	3	3	3	3	3
	ASGN	3	2	3	3	3	3
	+/-		- 1				

Table 4-11

PERSONNEL REQUIRED BY SKILL, BY NUMBER OF DOCKS, BY NUMBER OF SHIFTS

SHIFTS	3						2						1					
DOCKS	A	E	H	M	Q	N	A	E	H	M	Q	N	A	E	H	M	Q	N
4	6	12	6	6	3	3	4	8	4	4	2	2	2	4	2	2	1	1
3	6	12	6	6	3	3	4	8	4	4	2	2	2	8	2	2	1	1
2	6	6	3	6	3	3	4	4	2	4	2	2	2	2	1	2	1	1

Table 4-12

DAYS REQUIRED TO COMPLETE INSPECTION BY DOCKS, BY SHIFT, BY AIRCRAFT POSSESSED.

SHIFTS / DOCKS	3			2			1		
	72	48	36	72	48	36	72	48	36
4	6	4	3	9	6	4½	18	12	9
3	8	5 1/3	4	12	8	6	24	16	12
2	12	8	6	18	12	9	36	24	18

Table 4-13

REQUIRED VERSUS ASSIGNED BY SKILL

SKILL / BASE		COL (48) 4 X 2	LAU (36) 2 X 3	REE (72) 4 X 3	VAN (48) 4 X 2	WIL (72) 4 X 3	SHE (48) 4 X 2
APG	RQD	4	6	6	4	6	4
	ASGN	54	43	82	51	88	56
	+/-	+50	+39	+76	+47	+82	+52
ENG	RQD	8	6	12	8	12	8
	ASGN	9	7	10	10	11	9
	+/-	+ 1	+ 1	- 2	+ 2	- 1	+ 1
HYD	RQD	4	3	6	4	6	4
	ASGN	5	3	7	3	6	5
	+/-	+ 1		+ 1	- 1		+ 1
MECH	RQD	4	6	6	4	6	4
	ASGN	6	5	7	6	5	5
	+/-	+ 2	- 1	+ 1	+ 2	- 1	+ 1
QC	RQD	2	3	3	2	3	2
	ASGN	3	2	3	3	3	3
	+/-	+ 1	- 1		+ 1		+ 1
NDI	RQD	2	3	3	2	3	2
	ASGN	3	2	3	3	3	3
	+/-	+ 1	- 1		+ 1		+ 1

4.6 WORKLOAD FORECASTING

The supervisor of the Avionics Antenna Repair Shop, 36 TFW, at Bellis AFB is attempting to forecast his work load requirement for the coming month, July. Part of the shop's work load involves repairing antennas for the F-15. Since the introduction of the new F-15 squadron in January, the number of F-15 antenna Line Replaceable Units (LRUs) received for repair is as shown in Table 4-14, below. Forecast the F-15 antenna LRU work load for July.

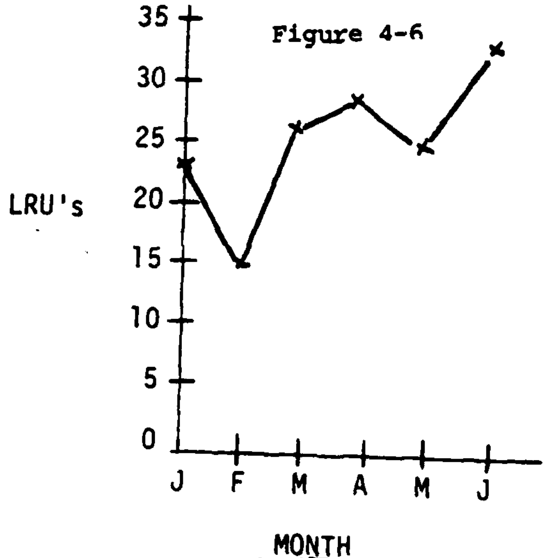
Table 4-14

F-15 ANTENNA LRU REPAIR REQUIREMENTS

JAN	FEB	MAR	APR	MAY	JUN
23	15	26	29	24	33

A graph of the data shown in Table 4-14 (see Figure 4-6, below) illustrates what appears to be an increasing trend in the number of antenna LRU's received for repair.

Figure 4-6



Number of Antenna LRU's Received for Repair by Month.

Using a simple moving average to forecast the July requirement would not reflect this trend. For example, a six-month moving average is:

$$\frac{23 + 15 + 26 + 29 + 24 + 33}{6} = 25$$

Where a system appears to be emerging and evolving, a better approach to forecasting is either exponential smoothing using relatively large values of α (the smoothing constant), or regression analysis.

In this problem, an exponentially smoothed forecast for July, using an α of .75 would be:

$$F_t = F_{t-1} + .75\ (33 - F_{t-1})$$

where F_{t-1} is the forecast for June.

If we had started with exponential smoothing in January, using 23 as both the forecast and actual value, the exponentially smoothed forecasts by month would be as shown in Table 4-15.

Table 4-15

EXPONENTIALLY SMOOTHED FORECASTS

JAN	FEB	MAR	APR	MAY	JUN
23*	23	17	23.8	27.7	24.9

*actual

For July then,

$$F_t = 24.9 + .75\ (33 - 24.9) = 30.975$$

Plotting the forecasts over the actual values in Figure 4-6 results in the picture shown in Figure 4-7. It can be seen that the exponentially smoothed forecast lags the actual. This is true of exponentially smoothed forecasts in general.

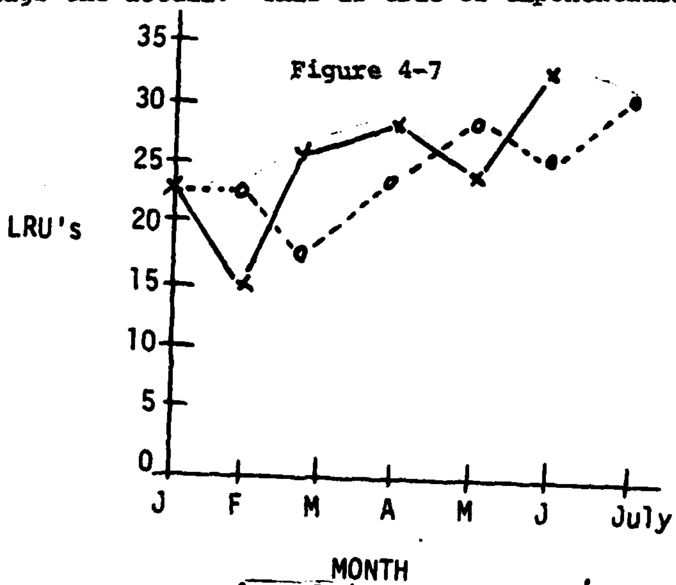


Figure 4-7

Actual (x) and Exponentially Smoothed (o) Values for LRU's

Using regression analysis, we can fit a straight line to the data using the least squares method. Computations are shown in Table 4-16.

Table 4-16

LEAST SQUARES COMPUTATION

LRU's Y	MONTH X	XY	X^2	Y^2
23	1	23	1	529
15	2	30	4	225
26	3	78	9	676
29	4	116	16	841
24	5	120	25	576
33	6	198	36	1089
150	21	565	91	3936

$$\bar{x} = \frac{21}{6} = 3.5 \qquad \bar{y} = \frac{150}{6} = 25.0$$

$$b = \frac{\sum xy - n\overline{xy}}{\sum x^2 - n\bar{x}^2} = \frac{565 - 6(3.5)(25.0)}{91 - 6(3.5)^2} = 2.286$$

$$a = \bar{y} - b\bar{x} = 25.0 - 2.286(3.5) = 17.0$$

Therefore

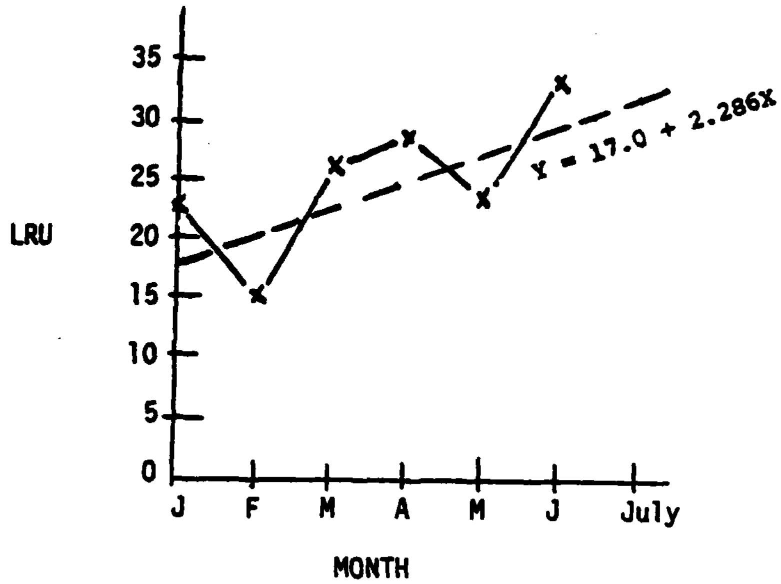
$$Y = a + bx = 17.0 + 2.286X$$

Using this model to predict the LRU work load requirement for July yields:

$$Y = a + bx = 17.0 + 2.286(7) = 33.002$$

The straight line plot is shown superimposed on the graph of actual values in Figure 4-8.

Figure 4-8



Least Squares Line and
Actual Values.

The Mean Absolute Deviation (MAD) for each of the forecasting methods, exponential smoothing (ES) and regression analysis (RA) can be computed as shown in Table 4-17. Based on this criterion, the forecast for July should be based on the regression model, i.e., 33. The disadvantage of using this method is the requirement to maintain all the data and the computational difficulty (although use of the computer or an advanced calculator eliminates the latter disadvantage).

Table 4-17

COMPUTATION OF MAD

		Forecast		Deviation	
Month	Actual	ES	RA	ES	RA
J (1)	23	--	--	--	--
F (2)	15	23	21.6	8	-6.6
M (3)	26	17	23.9	9	2.1
A (4)	29	23.8	26.1	5.2	2.9
M (5)	24	27.7	28.4	-3.7	-4.4
J (6)	33	24.9	30.7	8.1	2.3
J (7)	--	31.0	33.0	--	--
			$\lvert\sum\rvert$:	34.0	18.3

$$MAD_{ES} = \frac{\sum_{t=1}^{n} \left| A_t - F_t \right|}{n} = \frac{34.0}{5} = 6.8$$

$$MAD_{RA} = \frac{\sum_{t=1}^{n} \left| A_t - F_t \right|}{n} = \frac{18.3}{5} = 3.7$$

CHAPTER 5

TRANSPORTATION

This chapter contains several informational items that are quite often of particular concern to the transportation practitioner. As such, they are very pragmatic and mundane. The subject matter ranges from calculation of energy consumption by mode to simply finding the amount of liquid in a tank. However, these are the kinds of simple problems that usually occur at the first echelon of operations. Many of the items were taken directly from US Army Field Manual 55-15, Transportation Reference Data. Others were developed for this chapter. It is our firm hope that one or all may be of use to you as you function as a transportation practitioner.

5.1 LOGISTICS IN VIETNAM - HOW WE DID IT

Mail is essential to morale, especially in a combat zone. Unfortunately, the "official" evaluation of its importance as indicated by the transportation two priority assigned to it, is low. As a result there was usually a backlog of higher priority (i.e., a priority higher than two) cargo which should have taken movement precedence over mail, no matter how old. However, mail is a very sensitive matter to the soldier in the field and to his Congressman. Often, when a soldier or airman failed to get a letter at least every few days, his first thought was that the distribution system had failed. He would dash off a letter to his representative about the poor support being provided the poor soldier in the field. The legislator, aware of the clamor that can quickly grow from such cases, instituted a Congressional inquiry into the matter. Many of these landed on my desk. Usually everyone (soldier, airman, congressman, and military commander) was satisfied if it could be shown that at least some mail went to every unit, every day. However, mail was sacked and palletized by destination only. What was needed was some method that could "read" units out of destinations. The mechanism that was able to accomplish this seemingly impossible task was probability theory.

We were able to get data from the Postal Service which included the average percentages of mail shipped to each unit under given destination codes. Assuming a random process in the selection of sacks to go on a particular pallet for a given destination, we needed only to randomly select a sufficient number of sacks from pallets destined to a given location.

*NOTE: For the reader who wonders what ever happened to the nonselected mail sacks, please be consoled. After several days, mail became such a "critical" space taker at our terminal, extra airlift or surface transport was laid on to "relieve" the port.

These few bags were easily "stuffed" along with higher priority cargo, insuring near-daily deliveries to every unit. The number of congressional inquiries withered. Soldiers and airmen were reassured their interests were being supported. Commanders slept more easily. Statisticians sighed, "At last."

5.2 APPROXIMATING ENERGY CONSUMPTION BY MODE

Energy is, and will continue to be, a central topic in the world of transportation for many years to come. Yet, few transportation practitioners have directed their attentions to figuring out just how one might compute energy consumption by mode. One hears wild claims by many. A tractor-trailer rig at 65 miles per hour (MPH) gets better fuel mileage than the same rig at 55 MPH. A river barge beats all other modes in terms of energy consumption; therefore, we should encourage more canal building. Pipelines are the most energy efficient mode--they should be given the right of eminent domain. Sound familiar?

The following procedure will provide the practitioner a simple, programmed method of reducing each of the above claims to numbers. These numbers will not always be 100 percent accurate, but they are reasonable approximations of energy consumption under varying conditions for the several modes. They are particularly useful in evaluating competing modal claims and in discerning trends within a given mode. The accuracy of any of the estimations can be increased by substituting actual values whenever known for the average values supplied. In both the rail and the highway modes, resistance expressed in pounds per gross train or motor ton is related to several factors as shown below:

$$A + (B/W) + CV + DV^2$$

In the above expression, A is a fixed factor, B/W is a factor that varies by weight, CV represents the factor that varies with velocity, and DV^2 is a factor that varies with velocity squared. Each of the other modes has a similar set of factor relationships that determines its resistance. Obviously, it is very cumbersome to compute the resistance (R) for any or all of the modes even at a single speed. These calculations for the relevant velocity ranges in which we are most interested are graphically displayed on Figure 5-1.

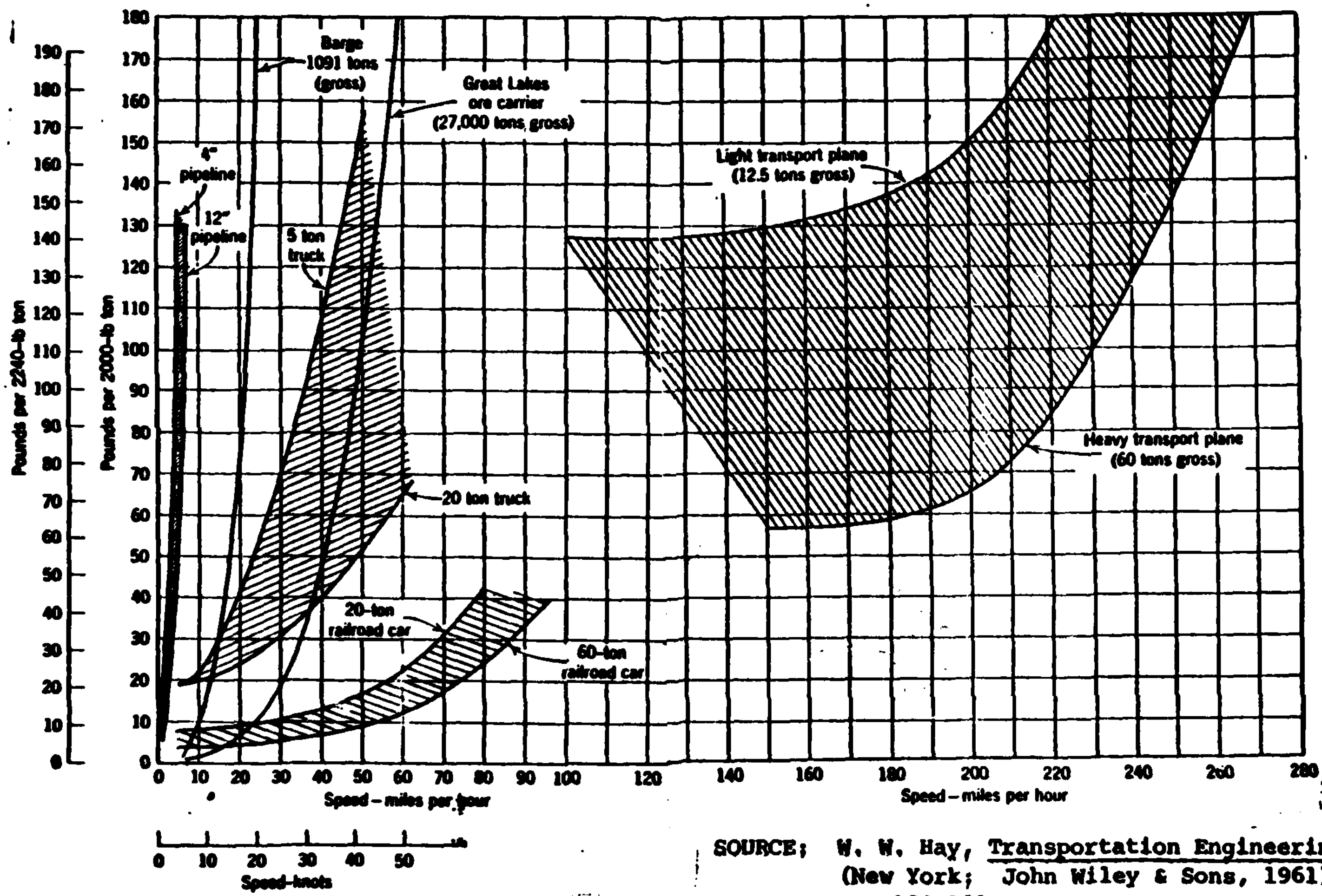
To get a value for the resistance in pounds per ton for any mode, simply read across the bottom scale to the desired velocity in miles per hour (MPH), then read up to the modal line. Read the value R from the vertical scale directly across (horizontally to the bottom scale) from the intersection of the velocity and the modal line.

<u>EXAMPLE 5-1</u>: Find R for a truck which weighs 5 tons gross, at 30 MPH.

1. Read across the bottom velocity scale to 30 MPH.
2. Read up the 30-MPH line to the 5-ton truck line.
3. Read the R value across from the velocity-mode line intersection.
4. The desired R value is 70 pounds per ton.

Table 5-1

Transport Technology Technological Characteristics



SOURCE: W. W. Hay, Transportation Engineering (New York; John Wiley & Sons, 1961): pp. 190-191.

EXAMPLE 5-2: Find R for a Great Lakes ore carrier of 27,000 tons gross at 30 MPH.

1. Read across bottom velocity scale to 30 MPH.
2. Read up the 30-MPH line to the 27,000-ton Great Lakes ore carrier line.
3. Read the R value across from the velocity-mode line intersection.
4. The desired R value is 19 pounds per ton.

Though the R values read from Figure 5-1 are a direct indication of just how much effort or energy must be exerted on a ton to keep it moving at any desired velocity, it is not expressed in the most common measure of energy intensity for the various modes. The most useful measure of energy efficiency is the British Terminal Unit (BTU) per ton-mile (TM).

All fuels have a known BTU content. One gallon of fuel oil, for example, contains approximately 130,000 BTUs. A pound of coal contains approximately 14,500 BTUs. Thus, no matter what fuel is consumed by a particular mode, the BTUs per ton-mile (TM) can be calculated, facilitating cross-modal comparisons.

To turn the R values of Figure 5-1 into BTU/TM, we first convert R into BTUs/hour, then divide this value by gross ton miles per hour.

BTU/hour = R <u>times</u> gross tons <u>times</u> velocity <u>times</u> 6.8

EXAMPLE 5-3: Find BTU/TM (gross) for a 5-ton truck at 30 MPH.

BTU/hour = 70 x 5 x 30 x 6.8

BTU/hour = 71,400

If a 5-ton vehicle is traveling at 30 MPH, it will generate 150 ton-miles (TM) per hour (30 MPH x 5 tons). Energy intensity for the 5-ton truck is 476 BTU/ton-miles (gross) (71,400 BTU/hour ÷ 150 TM/hour). The same calculations that we performed for the five-ton truck can be repeated for any mode, at any velocity. However, the resulting BTU/TM are still not directly comparable because of differing thermal efficiencies and percents of the gross weight dedicated to cargo as opposed to superstructure, frames, bed, etc.

Figure 5-2

Mode / Attribute	Rail	Water	Truck	Air	Bus	Auto	Pipe
% Full Factor	77	75	70	26	25	19	100
% Efficiency	24	21	25	13	25	25	25
Adj. Factor	.19	.16	.18	.03	.06	.05	.25

The BTU of the calculations so far are pure, absolute. We do not get 130,000 BTU from a gallon of fuel oil or gasoline that is put into our tanks. In fact, the thermal efficiencies of our modes range from 13 to 25 percent. (These are shown in Figure 5-2). To attain the 71,400 BTU/hour for the 5-ton truck, we will have to expend more BTU in raw fuel because the thermal efficiency of the average truck is only 25 percent. This requires an adjustment to BTU/TM calculation.

Similarly, the TM or ton-mile portion of the calculation requires some adjustment. A low BTU/TM is not very meaningful if most of the tons being hauled are devoted to the vehicle weight itself and not payload. This is the case with the airplane and the bus. Average percent full factors for each mode are shown in Figure 5-2. This percent represents that portion of the gross weight that is devoted to payload.

Fortunately, the adjustments for both the BTU and TM portions of the calculation can be made simultaneously using the adjustment factor shown in Figure 5-2. (This factor is the product of the percent full factor and the percent efficiency factor.) To correct BTU/TM (gross) to BTU/TM (net), simply divide BTU/TM (gross) by the adjustment factor.

<u>EXAMPLE 5-4:</u> Find the BTU/TM (net) of a 5-ton truck at 30 MPH.

1. R = 70 pounds per gross ton (from Figure 5-1)
2. BTU/hour = 70 x 5 x 30 x 6.8 = 71,400 BTU/hour
3. TM/hour = 5 x 30 = 150 TM/hour
4. BTU/TM (gross) = 71,400 ÷ 150 = 476 BTU/TM (gross)
5. BTU/TM (net) = 476 ÷ .18 = 2644.4

SOME APPLICATIONS

Let us reduce the slogans in the introduction of this section to hard numbers.

EXAMPLE 5-5: A large tractor-trailor rig (20 tons) gets better fuel mileage at 65 MPH than at 55 MPH.

Find BTU/TM (net) at 55:

1. R = 58

2. BTU/TM (gross)

   a. BTU/hour = 58 x 20 x 55 x 6.8 = 433840

   b. TM/hour = 20 x 55 = 1100

   BTU/TM (gross) = 433840 ÷ 1100 = 394.4

3. BTU/TM (net) = 394.4 ÷ .18 = 2191.1

Find BTU/TM (net) at 65

1. R = 75

2. BTU/TM (gross)

   a. BTU/hour = 75 x 20 x 65 x 6.8 = 663000

   b. TM/hour = 20 x 65 = 1300

   BTU/TM (gross) = 663000 ÷ 1300 = 510

3. BTU/TM (net) = 510 ÷ .18 = 2833.3

Conclusion: The required expenditure in pure energy will not allow better fuel mileage at 65 than at 55.

NOTE: If you desire to estimate the gallons of fuel per hour for this rig at the two speeds of 55 or 65, proceed as follows: First, divide the BTU/hour by the % efficiency factor. Second, divide the resulting value of step one by 130,000.

At 65 MPH: 663000 ÷ .25 = 2652000
2652000 ÷ 130000 = 20.4 gallons per hour

At 55 MPH: 433840 ÷ .25 = 1735360
1735360 ÷ 130000 = 13.3 gallons per hour

Conclusion: It takes more gallons of fuel per hour to drive 65 MPH than 55 MPH.

EXAMPLE 5-6: A river barge beats all other modes in terms of energy consumption.

Find BTU/TM (net) for a 1091-ton barge at 22 MPH.

1. R = 100

2. BTU/TM (gross)

   a. BTU/hour = 100 x 1091 x 22 x 6.8 = 16,321,360

   b. TM/hour = 1091 x 22 = 24002

   BTU/TM (gross) = 16321360/24002 = 680

3. BTU/TM = 680 ÷ .16 = 4250 BTU/TM

Conclusion: 4250 BTU/TM (net) > BTU/TM (net) of a 20-ton tractor-trailer @ 65 MPH. Barge is not most energy efficient carrier at all speeds.

EXAMPLE 5-7: A pipeline is the most energy efficient mode.

Find BTU/TM (net) for a 12" pipeline at 7 MPH. [NOTE: To determine ton miles for the pipeline, we must first convert the flow into tons. In one hour, seven miles of fluid will pass a given point. That is, the flow past a given point is equivalent to a column of fluid 36,960 (7 miles x 5280 feet/mile) feet tall and one foot in diameter. This is a volume of 29029.3 cubic feet (36960 feet x .7854 square feet of cross sectional area), or 743 tons, if the fluid is oil. (One cubic foot of oil weighs approximately 51.2 pounds.)]

1. R = 130

2. BTU/TM (gross)

   a. BTU/hour = 130 x 743 x 7 x 6.8 = 4,597,684

   b. BTU/hour = 743 x 7 = 5201

   BTU/TM (gross) = 4597684 ÷ 5201 = 884

3. BTU/TM (net) = 884 ÷ .25 = 3536

Conclusion: The 12-inch pipeline is not the most energy efficient mode at 7 MPH at the BTU/TM (net) > the BTU/TM (net) of a 20-ton tractor-trailer rig at 65 MPH as calculated above.

By reducing each of the slogans cited in the introduction of this energy approximation method to hard numbers, none could be supported at the speed values selected. Rhetoric is replaced by reason; emotion by evaluation. Admittedly, the simplified procedure is only an approximation. But, part of its worth beyond being a reasonable approximation lies in the ease of increasing its accuracy. If R, the fuel efficiency factor, or the percent full factor is known for a particular situation, any one or combination of the actual values may be substituted for the average values of figures one or two directly. The procedure requires no modification to accept these actual values.

<u>EXAMPLE 5-8:</u> Find the BTU/TM (net) of 60-ton railroad car at 55 MPH.

1. R = 10

2. BTU/TM (gross)

   a. BTU/hour = 10 x 60 x 55 x 6.8 = 224400

   b. TM/hour = 60 x 55 = 3300

   BTU/TM (gross) = 224400 ÷ 3300 = 68

3. BTU/TM (net) = 68 ÷ .19 = 358

Answer: 358 BTU/TM (net)

<u>EXAMPLE 5-9</u>: Find BTU/TM (net) of 60-ton heavy transport plane at 260 MPH.

1. R = 158

2. BTU/TM (gross)

   a. BTU/hour = 158 x 60 x 260 x 6.8 = 16,760,640

   b. TM/hour = 60 x 260 = 15600

   BTU/TM (gross) = 16760640 ÷ 15600 = 1074

3. BTU/TM (net) = 1074 ÷ .03 = 35813

Answer: 35813 BTU/TM (net)

[NOTE: The values for the transport planes of Figure 5-1 do not reflect jet aircraft. These aircraft average about 10000 BTU/TM (net). For specific values see: <u>Transportation Vehicle Energy Intensities</u>. US Dept of Transportation/NASA. Reference Paper, June 20, 1974.]

5.3 GRAPHICAL SCHEDULING TECHNIQUE

Frequently, the transportation practitioner is faced with the problem of dispatching vehicles or convoys of uniform makeup onto single lane rights-of way. This could occur on single width roadways under austere conditions or on single track rail lines.

As one vehicle after another is dispatched outbound, at some point in time they will begin encountering returning vehicles. Because they are on single vehicle width tracks or roadways, provisions for passing in the form of turn-outs, bypasses, or pull-around must be provided.

Tracking the movements for such a system can be a confusing and difficult problem. This is especially true when this activity occurs under the primitive conditions of the battlefield where sophisticated data processing and automatic recording equipment are not always available. The following Graphical Scheduling Technique (GST) provides a ready, hand-calculable approach to the problem.

On the graphical schedule (see Figure 5-1) each right descending diagonal line represents a departure from an origin station at mile 0. At mile 180, the vehicle, convoy, or train* immediately turns around and proceeds towards the origin station at mile 0. The time that any vehicle will pass a given point is read directly from the graph.

For example, the vehicle that departed mile zero at 12 midnight passed mile 60 at 6:00 a.m., mile 120 at 12 noon and arrived at mile 180 at 6:00 p.m. (These times and distances are marked with circles on Figure 5-3). Also, the location of a vehicle at any time, or conversely, the time at which a vehicle will pass a given location mile marker is easily found.

From the graph it can be seen that the midnight vehicle makes an immediate turnaround at mile marker 180, and starts back toward the origin. Extending the graph to seventy-two hours the schedule for the midnight departing vehicle indicates two complete round trips between mile zero and mile 180, with the second arrival at the origin occurring precisely at the same time of day as the original departure of midnight. (See Figure 5-4).

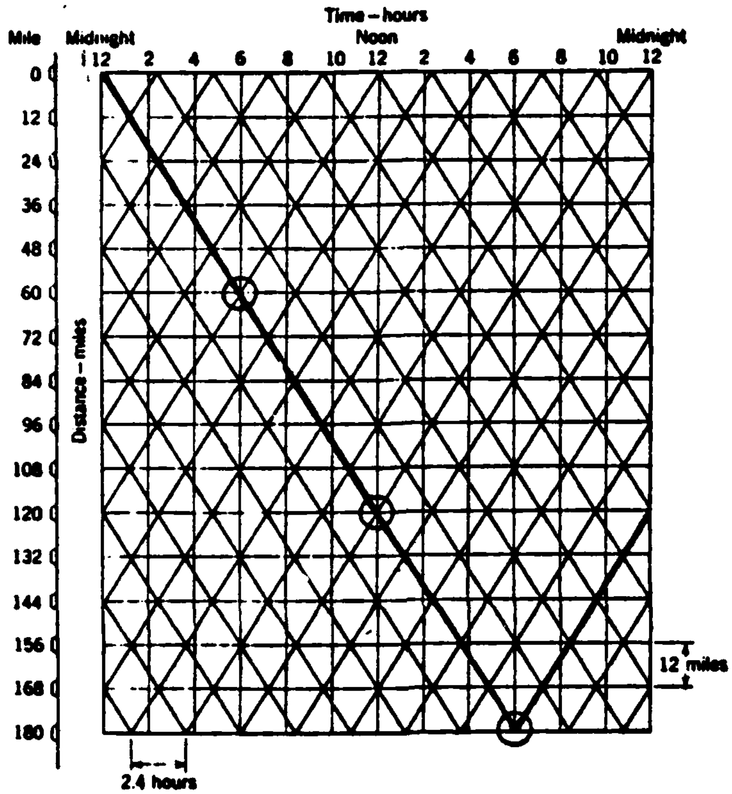
If the original midnight departing vehicle makes a complete cycle in 72 hours (3 days), it follows that the three-day cycle applies to any vehicle departing mile zero at any hour of the day, if it proceeds at the speed of the original vehicle. Thus the schedule shown in Figure 5-2 actually depicts an endless succession of three-day cycles of vehicles dispatched from mile zero to mile 180 and returning to mile zero.

A vehicle dispatched at midnight on day one of a given month will be ready to be dispatched again from mile zero at midnight on day 4.

---

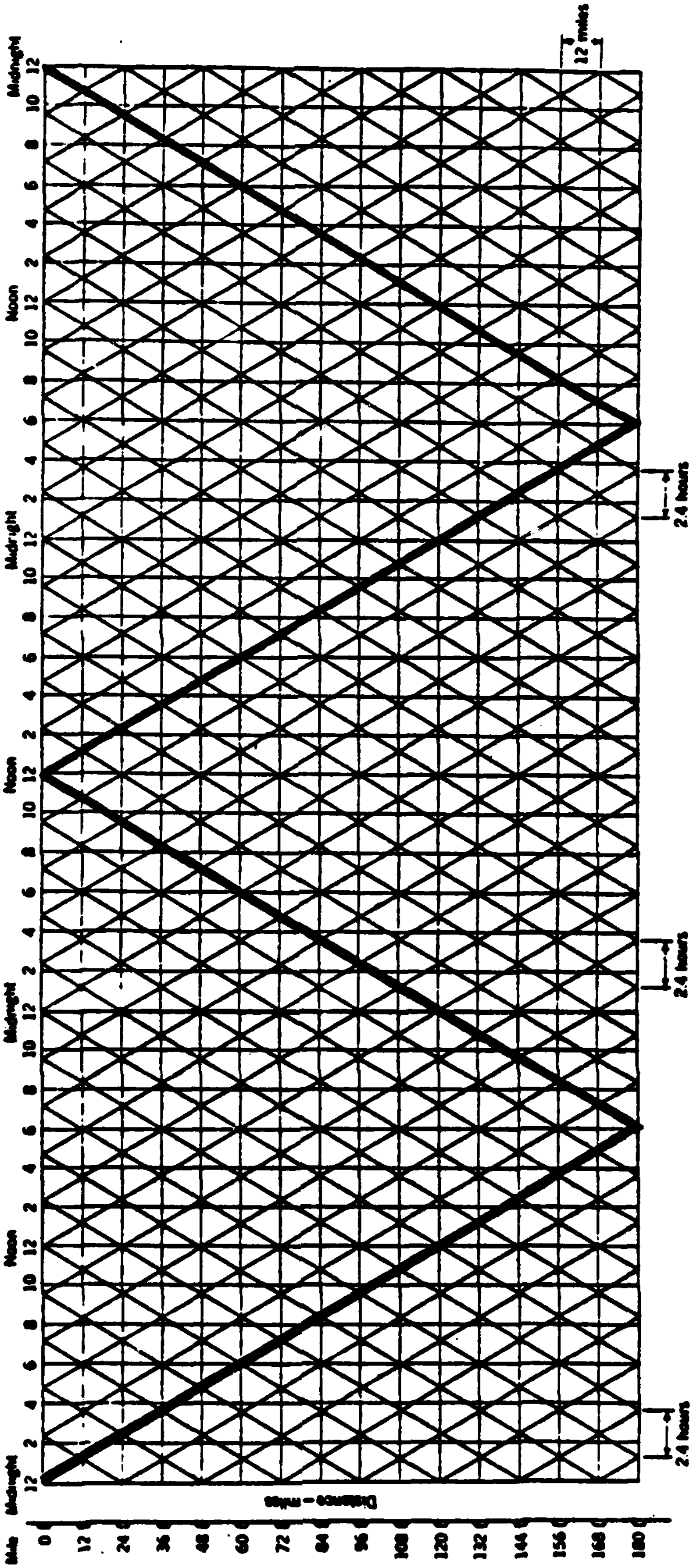
*To simplify terminology the rest of the discussion will use the single term vehicle to describe a vehicle, convoy, or train.

Figure 5-3



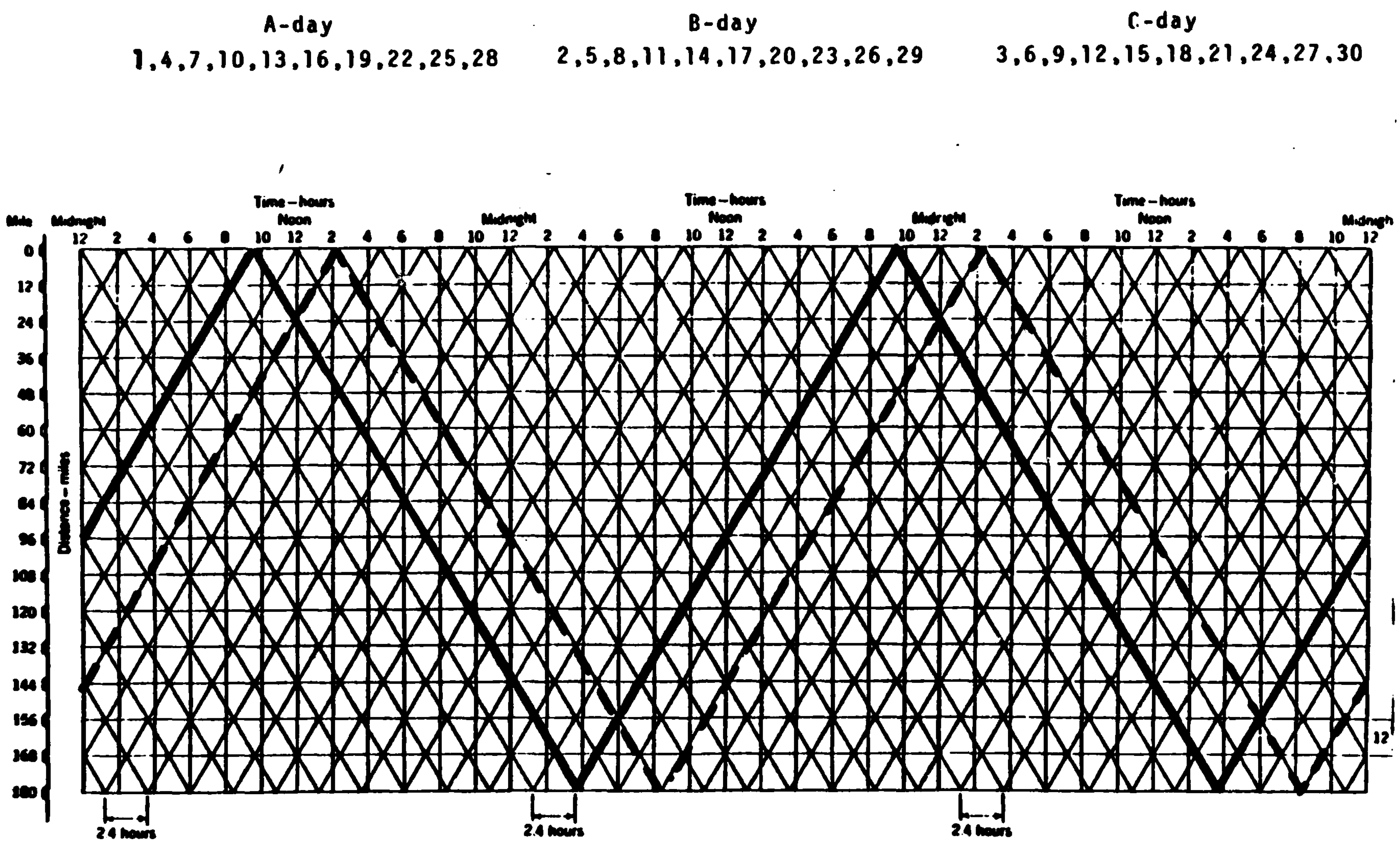
NOTE: For notational purposes the vehicle departing mile 0 at midnight is designated vehicle, train or convoy one, that which departs mile 0 at 0224 as number two, and so on.

Figure 5-4



Time - hours
Noon
Midnight
Distance - miles
Mile
0
12
24
36
48
60
72
84
96
108
120
132
144
156
168
180
2.4 hours
12 miles

Figure 5-5


A-day
1,4,7,10,13,16,19,22,25,28
B-day
2,5,8,11,14,17,20,23,26,29
C-day
3,6,9,12,15,18,21,24,27,30
Time-hours
Noon
Midnight
Mile
Distance - miles
0
12
24
36
48
60
72
84
96
108
120
132
144
156
168
180
2.4 hours
12

If dispatched immediately, at midnight of day four, it will be back at mile zero at midnight of day seven. This continues indefinitely. Thus day one for any month represents the system's status (assuming that vehicles are on time) on days 4, 7, 10, 13, 16, 19, 22, 25 and 28. Similarly, day 2 of the original three-day schedule represents days 5, 8, 11, 14, 17, 20, 23, 26, 29. Day 3 represents days 6, 9, 12, 15, 18, 21, 24, 27, 30. (The days that each 24-hour period represent are shown above each of the three 24-hour periods of Figure 5-5. For simplification, the first 24 hour period is labeled an A-day; the second, a B-day and the third, a C-day)

From Figure 5-5 one can read directly for the vehicle that departed mile zero at 0936 of day one (shown as a heavy, solid black line), that it will pass mile marker 84 on day 22 (or any of the other days of the month shown with day 1). On the 27th of the month at 1800, it will be at mile marker 156, traveling towards the mile marker zero. It will have arrived at mile marker 156 at the same time as the vehicle that left mile marker zero 48 hours behind it, on day 1.

GST provides a method of determining where each vehicle will be at any point in time during any desired period. (The period is not limited to a month. The use of the Julian calendar would easily extend the period to a full year where, if Day A represented 1 January, or 001 on the Julian calendar, Day C would represent Julian day 366 in a leap year.) It also dictates where each bypass, turnout, or pull around must be located. It indicates vehicle spacing, an item that affects safety. For example, from Figure 5-5 at 6 a.m. on day 7 (Day A), train 3 is at mile 12, train 2 at mile 36. Trains 1, 15, 14, 13, 12, and 11 are at mile markers 60, 84, 108, 132, 156, and 180 respectively. Spacings are a constant 24 miles.

In many situations the basic GST method explained above can be very helpful. However, its flexibility to accommodate more complex situations is deceiving.

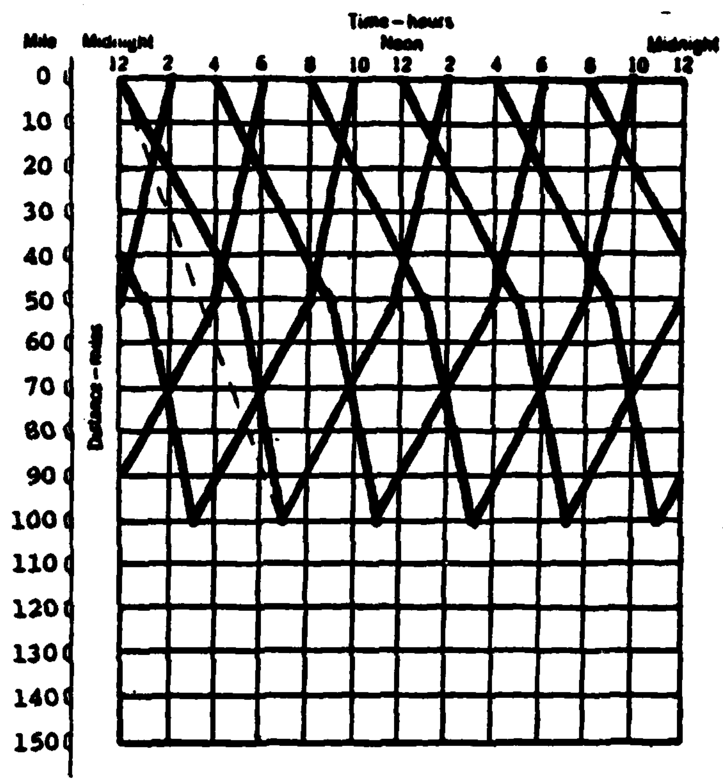
That the slope of the diagonals represents average speed of the train is not immediately apparent. The slope of the diagonal for train one is 180 miles divided by 18 hours or 10 miles per hour. (A positive slope indicates outbound movement). Thus, GST can accommodate any speed for any particular distance segment simply by changing the slopes of the diagonal for that segment.

Figure 5-6 depicts a two-speed situation. Between mile marker 0 and 50, the train averages 10 mph. Between 50 and 100, it averages 25 mph. The overall average, represented by the dashed line, is 14.29 mph.

Assuming an hour turnaround at the end of each roundtrip at mile zero, four trains are required to make a full schedule with a departure every four hours.

The completed graphic schedule indicates that four pull bys or bypasses are required. One each at miles 16, 44, and 72; one at the turnaround point, mile 100.

Figure 5-6



Time – hours
Mile
Midnight
Noon
Midnight
12
2
4
6
8
10
12
2
4
6
8
10
12
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
Distance – miles

Figure 5-6 also represents a common situation where the terrain might be predominantly uphill to mile 50, and downhill from 50 to 100. The return trip is similarly uphill for the first half and downhill over the second. This accounts for the different speeds over the two halves in both directions. In summary, GST is a simple scheduling technique that may be used in primitive conditions. It indicates where all vehicles are located at any time. It can also be used to predict where any vehicle will be at some future time. It is flexible enough to accommodate a variety of speeds over the given course. Perhaps most important, it simplifies the locating of those spots where vehicles must pass.

5.4 VOLUME OF LIQUID IN TANKS

a. General. Transportation planners and users frequently need weight and cube data for liquids. After volume of liquid has been determined, the weight may be found by multiplying the total volume in gallons by the weight of 1 gallon of the liquid. (Unit weights for various liquids are to be found in Miscellaneous Data.) It is assumed that the cylindrical and elliptical tanks discussed below have flat ends. (If a tank does not have flat ends, then the curvature of the ends must be determined, and the volume contained in the ends added to the volumes shown below. However, in such cases, it is usually more advantageous to calibrate the tank, which entails the drawing off or adding of measured amounts of the liquid and measuring the change in liquid depth in the tank). In the charts below, single and/or double interpolation should be used to determine values between those shown. Values shown are close approximations for 60° F. and can be used for practical purposes. Variations in conditions, especially temperature, cause the values to vary. If exact quantities are desired, the formulas shown must be used in conjunction with temperature coefficients of expansion and contraction for the liquids and containers. Presentation of detailed computations for varying conditions is not within the scope of this manual, but approximate computations for POL products under various temperatures are shown in f below. Unless otherwise specified, dimensions are in inches.

b. Cylindrical Tanks

(1) Vertical

(a) Formula. The volume of liquid in a vertical cylindrical tank may be computed by using the formula shown below.

Gallons per foot of liquid depth

$$= \frac{(3.1416) \times (\text{radius of tank})^2 \times (12)}{231}$$

$$= \frac{(9.4248) \times (\text{diameter of tank})^2}{231}$$

$$= (.0408) \times (\text{diameter})^2$$

(*b*) *Table*. Gallons per foot of liquid depth are shown for vertical cylindrical tanks of various diameters.

Diameter (in.)	*Gallons per ft*	*Diameter (in.)*	*Gallons per ft*	*Diameter (in.)*	*Gallons per ft*
1	0.0408	54	118.9	107	467.5
2	0.163	55	123.3	108	477.5
3	0.367	56	127.9	109	486.0
4	0.653	57	132.8	110	496.0
5	1.02	58	137.4	111	503.0
6	1.47	59	142.2	112	512.5
7	2.00	60	147.1	113	522.0
8	2.61	61	152.1	114	532.5
9	3.30	62	157.0	115	541.0
10	4.08	63	162.0	116	551.0
11	4.93	64	167.5	117	560.0
12	5.87	65	172.5	118	570.0
13	6.89	66	177.7	119	580.0
14	8.00	67	183.3	120	589.0
15	9.17	68	188.3	121	597.0
16	10.44	69	194.5	122	607.0
17	11.78	70	200.0	123	617.0
18	13.22	71	206.0	124	627.0
19	14.72	72	211.9	125	637.0
20	16.31	73	217.5	126	648.0
21	18.00	74	223.6	127	658.0
22	19.75	75	229.8	128	668.0
23	21.60	76	236.0	129	679.0
24	23.50	77	242.2	130	689.0
25	25.50	78	248.7	131	698.0
26	27.56	79	255.0	132	710.0
27	29.79	80	261.2	133	722.0
28	32.00	81	268.0	134	733.0
29	34.33	82	274.8	135	743.0
30	36.65	83	281.0	136	754.0
31	39.20	84	288.0	137	766.0
32	41.80	85	295.3	138	777.0
33	44.40	86	302.0	139	788.0
34	47.20	87	309.0	140	800.1
35	49.90	88	316.8	141	811.0
36	52.80	89	323.5	142	822.0
37	55.80	90	331.0	143	833.0
38	58.90	91	338.0	144	846.0
39	62.00	92	346.0	145	858.0
40	65.30	93	353.6	146	871.0
41	68.60	94	361.0	147	882.0
42	71.90	95	369.0	148	893.0
43	75.40	96	377.0	149	906.0
44	79.00	97	384.5	150	917.0
45	82.60	98	392.5	151	930.0
46	86.30	99	401.0	152	942.0
47	90.10	100	408.5	153	955.0
48	93.90	101	417.0	154	968.0
49	97.90	102	426.0	155	979.0
50	102.0	103	434.0	156	993.0
51	106.0	104	442.5	157	1,006
52	110.1	105	451.0	158	1,018
53	114.6	106	459.0	159	1,031

Diameter (in.)	Gallons per ft	Diameter (in.)	Gallons per ft	Diameter (in.)	Gallons per ft
160	1,044	218	1,940	276	3,110
161	1,058	219	1,957	277	3,125
162	1,070	220	1,975	278	3,150
163	1,082	221	1,991	279	3,180
164	1,096	222	2,010	280	3,200
165	1,110	223	2,025	281	3,220
166	1,122	224	2,050	282	3,244
167	1,137	225	2,065	283	3,265
168	1,151	226	2,083	284	3,288
169	1,165	227	2,100	285	3,310
170	1,178	228	2,120	286	3,335
171	1,193	229	2,140	287	3,360
172	1,205	230	2,160	288	3,382
173	1,219	231	2,179	289	3,408
174	1,235	232	2,193	290	3,433
175	1,247	233	2,212	291	3,455
176	1,263	234	2,233	292	3,475
177	1,278	235	2,250	293	3,498
178	1,293	236	2,270	294	3,529
179	1,309	237	2,291	295	3,550
180	1,322	238	2,310	296	3,572
181	1,336	239	2,330	297	3,595
182	1,347	240	2,350	298	3,620
183	1,365	241	2,370	299	3,643
184	1,380	242	2,389	300	3,665
185	1,395	243	2,409	301	3,690
186	1,411	244	2,429	302	3,720
187	1,425	245	2,449	303	3,743
188	1,442	246	2,469	304	3,770
189	1,459	247	2,489	305	3,800
190	1,472	248	2,509	306	3,820
191	1,486	249	2,530	307	3,850
192	1,504	250	2,550	308	3,880
193	1,520	251	2,571	309	3,900
194	1,537	252	2,590	310	3,920
195	1,553	253	2,610	311	3,950
196	1,566	254	2,630	312	3,970
197	1,580	255	2,650	313	4,000
198	1,600	256	2,672	314	4,030
199	1,618	257	2,694	315	4,050
200	1,631	258	2,717	316	4,080
201	1,648	259	2,734	317	4,110
202	1,666	260	2,754	318	4,130
203	1,680	261	2,780	319	4,160
204	1,697	262	2,800	320	4,190
205	1,714	263	2,819	321	4,210
206	1,733	264	2,844	322	4,230
207	1,748	265	2,862	323	4,260
208	1,766	266	2,890	324	4,290
209	1,781	267	2,909	325	4,310
210	1,800	268	2,930	326	4,340
211	1,819	269	2,950	327	4,360
212	1,836	270	2,979	328	4,390
213	1,851	271	2,992	329	4,420
214	1,870	272	3,018	330	4,440
215	1,884	273	3,029	331	4,470
216	1,902	274	3,061	332	4,500
217	1,920	275	3,085	333	4,530

Diameter (in.)	Gallons per ft	Diameter (in.)	Gallons per ft	Diameter (in.)	Gallons per ft
334	4,550	455	8,450	770	24,220
335	4,580	460	8,630	780	24,870
336	4,610	465	8,830	790	25,500
337	4,630	470	9,010	800	26,120
338	4,670	475	9,200	810	26,800
339	4,690	480	9,390	820	27,480
340	4,720	485	9,600	830	28,100
341	4,740	400	9,790	840	28,800
342	4,770	495	10,000	850	29,530
343	4,800	500	10,200	860	30,200
344	4,830	505	10,400	870	30,900
345	4,860	510	10,600	880	31,680
346	4,890	515	10,810	890	32,350
347	4,920	520	11,010	900	33,100
348	4,950	525	11,240	910	33,800
349	4,980	530	11,460	920	34,600
350	4,990	535	11,670	930	35,360
351	5,030	540	11,890	940	36,100
352	5,060	545	12,100	950	36,900
353	5,080	550	12,330	960	37,700
354	5,110	555	12,560	970	38,450
355	5,150	560	12,790	980	39,250
356	5,180	565	13,090	990	40,100
357	5,210	570	13,280	1,000	40,850
358	5,240	575	13,500	1,010	41,700
359	5,270	580	13,750	1,020	42,600
360	5,290	585	13,980	1,030	43,400
365	5,440	590	14,220	1,040	44,250
370	5,580	600	14,700	1,050	45,100
375	5,740	610	15,210	1,060	45,900
380	5,890	620	15,700	1,070	46,750
385	6,060	630	16,200	1,080	47,750
390	6,200	640	16,750	1,090	48,600
395	6,370	650	17,250	1,100	49,500
400	6,530	660	17,770	1,110	50,300
405	6,700	670	18,330	1,120	51,250
410	6,860	680	18,880	1,130	52,200
415	7,030	690	19,450	1,140	53,250
420	7,190	700	20,000	1,150	54,100
425	7,380	710	20,600	1,160	55,100
430	7,540	720	21,190	1,170	56,000
435	7,730	730	21,750	1,180	57,000
440	7,900	740	22,360	1,190	58,000
445	8,080	750	22,980	1,200	58,900
450	8,260	760	23,600		

(2) *Horizontal.*

(*a*) When the capacity of a horizontal cylindrical tank is known, the quantity in the tank may be approximated by using the scale below.

Part of tank depth filled	*Part of tank capacity filled*
	1.000
0.95	.974
.90	.948
.85	.904
.80	.860
.75	.804
.70	.740
.65	.687
.60	.626
.55	.563
.50	.500
.45	.437
.40	.374
.35	.313
.30	.252

Part of tank depth filled	*Part of tank capacity filled*
.25	.196
.20	.140
.15	.096
.10	.052
.05	.026

Example. A horizontal tank is 80 inches in diameter. Depth of liquid in tank is 20 inches. Full tank capacity is 8,000 gallons. Find the number of gallons actually in tank.

$\frac{20}{80} = 0.25$ (part of tank depth filled)

0.25 (part of tank depth filled) = 0.196 (part of tank capacity filled)

$0.196 \times 8{,}000 = 1{,}568$ gallons in tank

(*b*) When the capacity of a horizontal cylindrical tank is not known, the volume of liquid may be computed as shown below.

L = length of tank
l = depth of liquid in tank
r = radius of tank
h = distance from top of tank to surface of liquid
D = diameter of tank
a = cross-sectional length of "wet arc" formed by liquid, measured on lower part of tank
a' = cross-sectional length of "dry arc" above liquid, measured on upper part of tank

1. When tank is less than half full:

$$\text{Volume (gal.)} = \left[\frac{ar}{2} - (r - l)\ \sqrt{2rl - l^2}\right] \times \frac{L}{231}$$

2. When tank is half full:

$$\text{Volume (gal.)} = \frac{r^2L}{147}$$

3. When tank is more than half full:

$$\text{Volume (gal.)} = \left[3.1416r^2 - a'r + (r - h)\ \sqrt{2rh - h^2}\right] \times \frac{L}{231}$$

(*c*) When it is not practical to measure a and a' in the formulas above (because of buried tanks, etc.), the lengths of these arcs can be computed by determining l/D and h/D ratios, and using the table below.

Tank less than half full

l/D	*Quantity to multiply D by to obtain a*	*l/D*	*Quantity to multiply D by to obtain a*
.01	0.200	.26	1.070
.02	0.284	.27	1.093
.03	0.348	.28	1.115
.04	0.403	.29	1.137
.05	0.451	.30	1.159
.06	0.495	.31	1.181
.07	0.536	.32	1.203
.08	0.574	.33	1.224
.09	0.609	.34	1.245
.10	0.644	.35	1.266
.11	0.676	.36	1.287
.12	0.708	.37	1.308
.13	0.738	.38	1.328
.14	0.767	.39	1.349
.15	0.795	.40	1.369
.16	0.823	.41	1.390
.17	0.850	.42	1.410
.18	0.876	.43	1.430
.19	0.902	.44	1.451
.20	0.927	.45	1.471
.21	0.952	.46	1.491
.22	0.976	.47	1.511
.23	1.000	.48	1.531
.24	1.024	.49	1.551
.25	1.047	.50	1.571

Tank more than half full

h/D	*Quantity to multiply D by to obtain a'*	*h/D*	*Quantity to multiply D by to obtain a'*
.01	0.200	.26	1.070
.02	0.284	.27	1.093
.03	0.348	.28	1.115
.04	0.403	.29	1.137
.05	0.451	.30	1.159
.06	0.495	.31	1.181
.07	0.536	.32	1.203
.08	0.574	.33	1.224
.09	0.609	.34	1.245
.10	0.644	.35	1.266
.11	0.676	.36	1.287
.12	0.708	.37	1.308
.13	0.738	.38	1.328
.14	0.767	.39	1.349
.15	0.795	.40	1.369
.16	0.823	.41	1.390
.17	0.850	.42	1.410
.18	0.876	.43	1.430
.19	0.902	.44	1.451
.20	0.927	.45	1.471
.21	0.952	.46	1.491
.22	0.976	.47	1.511
.23	1.000	.48	1.531
.24	1.024	.49	1.551
.25	1.047	.50	1.571

(*d*) The volume of liquid in flat-end horizontal cylindrical tanks of various sizes is given below. Quantities are in U.S. gallons for each liquid depth shown per foot of tank length. Therefore, in the formulas shown in (*b*) above, the ratio $L/231$ becomes a constant of 0.0519. To obtain liquid volume, multiply the figure from the chart by the tank length (feet and fractions).

Depth of liquid in tank (in.)

Tank diameter (in.)	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60
9	.95	2.34	3.30																	
12	1.15	2.95	4.73	5.9																
15	1.30	3.41	5.67	7.6	9.18															
18	1.45	3.85	6.61	9.4	11.8	13.2														
21	1.58	4.24	7.37	10.5	13.5	16.0	18.0													
24	1.70	4.61	8.10	11.8	15.5	18.9	21.8	23.5												
27	1.78	4.86	8.70	12.8	17.0	21.0	24.7	27.5	29.8											
30	1.91	5.23	9.30	13.7	18.4	23.0	27.5	31.5	34.8	36.8										
33	1.99	5.43	9.80	14.5	19.6	24.8	29.8	34.5	38.7	41.9	44.4									
36	2.12	5.80	10.3	15.4	20.9	26.4	32.0	37.5	42.5	47.1	50.8	52.9								
39	2.20	5.94	10.8	16.2	22.0	28.0	34.0	40.0	45.7	51.1	55.8	59.3	62.2							
42	2.28	6.31	11.3	17.0	23.1	29.5	36.0	42.5	48.9	55.0	60.7	65.7	69.7	72.0						
45	2.35	6.43	11.8	17.7	24.1	30.9	37.8	44.8	51.7	58.4	64.8	70.7	75.9	79.6	82.7					
48	2.45	6.78	12.2	18.4	25.1	32.2	39.5	47.0	54.5	61.8	68.9	75.6	81.8	87.2	91.6	94.0				
51	2.54	6.89	12.6	19.1	26.1	33.5	41.2	49.0	57.0	64.8	72.6	80.0	86.9	93.3	98.8	102.9	106.3			
54	2.61	7.23	13.0	19.7	27.0	34.7	42.8	51.1	59.5	67.9	76.2	84.3	92.0	99.3	105.9	111.8	116.4	119.0		
57	2.68	7.43	13.4	20.3	27.9	35.9	44.1	53.0	61.8	70.7	79.5	88.2	96.6	104.6	112.1	118.9	124.8	129.2	132.8	
60	2.76	7.64	13.8	20.9	28.7	37.1	45.8	54.9	64.1	73.5	82.8	92.0	101.1	109.8	118.2	126.0	133.1	139.3	144.1	146.9
63	2.82	7.84	14.2	21.5	29.5	38.2	47.3	56.7	66.3	76.0	85.9	95.5	105.2	114.6	123.7	132.3	140.2	147.5	153.7	158.3
66	2.89	8.05	14.6	22.1	30.4	39.3	48.7	58.4	68.4	78.6	88.9	99.1	109.3	119.3	129.1	138.5	147.4	155.7	163.2	169.7
69	2.95	8.23	14.9	22.5	31.2	40.2	50.0	60.1	70.5	81.0	91.7	102.6	113.2	123.7	134.1	144.2	153.8	163.0	171.4	179.1
72	3.02	8.42	15.3	23.2	31.9	41.4	51.3	61.7	72.5	83.4	94.5	105.8	117.0	128.1	139.1	149.8	160.2	170.2	179.6	188.4
75	3.08	8.60	15.6	23.5	32.7	42.4	52.6	63.3	74.4	85.7	97.2	108.9	120.5	132.2	143.7	155.1	166.1	176.8	187.0	196.7
78	3.15	8.78	15.9	24.2	33.4	43.3	53.9	64.9	76.3	88.0	99.9	112.0	124.1	136.3	148.3	160.3	172.0	183.4	194.4	204.9
81	3.20	8.95	16.2	24.7	34.2	44.3	55.1	66.5	78.1	90.2	102.4	115.0	127.5	140.2	152.7	165.2	177.5	189.5	201.2	212.4
84	3.25	9.12	16.6	25.2	34.9	45.2	56.3	67.9	79.9	92.3	105.0	117.9	130.9	144.0	157.0	170.0	182.9	195.6	208.0	220.0
87	3.34	9.29	16.9	25.7	35.6	46.2	57.5	69.4	81.7	94.3	107.4	120.7	134.1	147.6	161.2	174.7	188.0	201.3	214.4	227.0
90	3.43	9.45	17.2	26.2	36.2	47.1	58.6	70.8	83.4	96.4	109.8	123.5	137.3	151.2	165.3	179.3	193.2	207.0	220.7	234.1
93	3.47	9.61	17.5	26.7	36.9	48.0	59.7	72.2	85.1	98.4	111.6	126.1	140.3	154.7	169.2	183.7	198.1	212.5	226.7	240.6
96	3.50	9.78	17.8	27.1	37.5	48.8	60.8	73.5	86.7	100.4	114.4	128.8	143.4	158.2	173.1	188.0	203.0	217.9	232.6	247.2
99	3.56	9.89	18.1	27.6	38.3	49.7	61.9	74.8	88.3	102.3	116.7	131.4	146.4	161.5	176.8	192.2	207.7	223.0	238.3	253.5
102	3.61	10.1	18.4	28.0	39.0	50.5	63.0	76.2	89.9	104.2	118.9	133.9	149.3	164.8	180.5	196.4	212.3	228.1	244.0	259.7
105	3.66	10.3	18.6	28.5	39.5	51.3	64.0	77.5	91.5	106.1	121.1	136.4	152.1	168.0	184.1	200.4	216.7	233.1	249.4	265.6
108	3.71	10.4	18.9	28.9	40.0	52.1	65.1	78.7	93.0	107.9	123.2	138.9	154.9	171.2	187.7	204.4	221.1	238.0	254.8	271.5
111	3.77	10.6	19.2	29.4	40.6	52.9	66.1	79.9	94.5	109.6	125.2	141.3	157.6	174.3	191.2	208.2	225.4	242.9	259.9	277.0
114	3.82	10.7	19.5	29.8	41.1	53.7	67.1	81.2	96.0	111.4	127.3	143.7	160.4	177.4	194.7	212.1	229.7	247.8	265.0	283.3
117	3.87	10.9	19.7	30.2	41.6	54.0	67.9	82.4	97.6	113.2	129.5	145.8	162.9	180.3	198.8	215.8	233.8	252.3	270.0	288.5
120	3.92	11.0	19.8	30.6	42.0	54.2	68.7	83.6	99.2	114.9	131.8	148.0	165.5	183.2	203.0	219.5	238.0	256.8	274.9	294.0

Tank diameter (in.)	Depth of liquid in tank (in.) 63	66	69	72	75	78	81	84	87	90	93	96	99	102	105	108	111	114	117	120
9																				
12																				
15																				
18																				
21																				
24																				
27																				
30																				
33																				
36																				
39																				
42																				
45																				
48																				
51																				
54																				
57																				
60																				
63	162.2																			
66	174.8	177.7																		
69	185.6	190.4	194.6																	
72	196.3	203.1	208.5	211.5																
75	205.6	213.6	220.4	225.5	230.0															
78	214.8	224.0	232.3	239.5	245.1	248.2														
81	223.2	233.4	242.7	251.1	258.2	263.5	268.0													
84	231.6	242,7	253.1	262.7	271.3	278.8	284.6	287.9												
87	239.4	251.2	262.5	273.1	282.8	291.6	299.0	303.9	308.7											
90	247.1	259.7	271.9	283.4	294.3	304.3	313.3	321.0	327.1	330.5										
93	254.4	267.7	280.6	292.9	304.8	315.8	327.9	335.0	342.6	343.4	352.9									
96	261.6	275.6	289.3	302.5	315.2	327.2	338.5	348.9	358.2	366.2	372.5	376.0								
99	268.4	283.1	297.6	311.4	324.8	337.8	350.0	361.4	371.9	381.4	389.3	395.2	399.5							
102	275.2	290.6	305.6	320.3	334.5	348.3	361.5	374.0	385.5	396.5	406.1	414.4	420.1	424.5						
105	281.7	297.6	313.3	328.7	343.6	358.2	372.2	385.6	398.2	410.1	421.0	430.7	437.1	445.0	449.5					
108	288.2	304.7	321.0	337.0	352.7	368.0	382.9	397.2	410.8	423.7	435.9	447.0	454.0	465.5	470.0	475.9				
111	294.4	311.5	328.4	345.0	361.3	377.4	392.9	407.6	422.4	436.4	449.0	461.5	471.0	482.8	490.5	497.4	502.6			
114	300.6	318.3	335.8	353.0	369.9	386.7	402.9	418.0	434.0	449.0	462.0	476.0	488.0	500.0	511.0	519.0	526.0	531.0		
117	306.3	324.2	342.4	360.0	377.9	394.9	411.9	428.3	444.0	460.5	474.8	489.8	503.0	516.0	527.0	537.5	546.5	553.6	558.0	
120	312.0	330.0	349.0	367.0	386.0	403.0	421.0	438.5	454.0	472.0	487.5	503.5	518.0	532.0	543.0	556.0	567.0	576.2	580.0	588.0

c. Rectangular Tanks. The chart below gives liquid volumes for each foot of liquid depth in rectangular tanks of various sizes. To obtain total liquid volume, multiply the figure shown by liquid depth (feet and fractions). Figures in the chart were obtained from the following formula:

$$\text{Volume (gal.)} = \frac{Lwl}{231}$$

Where

L = length of tank in inches
w = width of tank in inches
l = depth of liquid in inches

Width of tank (in.)	Length of tank (in.) 12	15	18	21	24	27	30	33	36
12	7.48	9.35	11.23	13.10	14.95	16.85	18.70	20.58	22.46
15	9.36	11.68	14.02	16.36	18.72	21.00	23.33	25.70	28.02
18	11.23	14.02	16.83	19.65	22.44	25.25	28.05	30.90	33.66
21	13.10	16.36	19.65	22.90	26.18	29.50	32.70	36.00	39.30
24	14.95	18.72	22.44	26.18	29.97	33.70	37.40	41.20	44.80
27	16.85	21.00	25.25	29.50	33.70	37.90	42.10	46.30	50.50
30	18.70	23.33	28.05	32.70	37.40	42.10	46.80	51.50	56.10
33	20.58	25.70	30.90	36.00	41.20	46.30	51.50	56.60	61.80
36	22.46	28.02	33.66	39.30	44.80	50.50	56.10	61.80	67.30
39	24.30	30.40	36.50	42.60	48.70	54.70	60.80	66.90	72.90
42	26.19	32.70	39.24	45.80	52.30	58.90	65.50	72.00	78.50
45	28.05	35.10	42.10	49.10	56.10	63.10	70.20	77.20	84.20
48	29.93	37.40	44.90	52.30	59.80	67.30	74.80	82.30	89.80
51	31.75	39.71	47.70	55.70	63.60	71.50	79.60	87.50	95.40
54	33.70	42.10	50.50	58.80	67.30	75.70	84.10	92.60	101.00
57	35.50	44.40	53.25	62.20	71.00	80.00	88.80	97.70	106.60
60	37.40	46.75	56.10	65.40	74.80	84.10	93.50	102.90	112.20
63	39.30	49.10	58.90	68.70	78.40	88.40	98.20	108.00	117.70
66	41.10	51.40	61.70	72.00	82.30	92.50	102.80	113.30	123.40
69	43.10	53.70	64.50	75.20	85.90	96.80	107.40	118.40	129.00
72	44.90	56.10	67.30	78.50	89.70	100.90	112.20	123.50	134.60
75	46.75	58.40	70.10	81.80	93.40	105.10	117.00	128.60	140.10
78	48.60	60.75	72.90	85.10	97.20	109.30	121.60	133.70	145.90
81	50.50	63.10	75.70	88.30	100.80	113.50	126.30	139.00	151.40
84	52.40	65.50	78.60	91.60	104.60	117.60	131.00	144.00	157.10
87	54.20	67.80	81.30	94.80	108.30	122.00	135.60	149.20	162.60
90	56.10	70.20	84.20	98.20	112.20	126.10	140.30	154.50	168.30
93	58.00	72.45	86.90	101.40	115.70	130.40	145.00	159.50	174.00
96	59.80	74.80	89.80	104.50	119.60	134.50	149.60	164.50	179.50
99	61.70	77.20	92.60	108.00	123.50	138.90	154.40	169.70	185.20
102	63.60	79.60	95.50	111.20	127.20	143.00	158.90	174.50	190.70
105	65.40	81.90	98.30	114.50	131.00	147.30	163.60	179.60	196.40
108	67.30	84.20	101.20	117.80	134.60	151.50	168.30	184.90	202.00
111	69.20	86.60	103.90	121.00	138.30	155.50	173.00	190.00	206.20
114	71.00	88.90	106.70	124.40	142.10	159.80	177.70	195.00	213.20
117	72.80	91.30	109.50	127.50	145.90	164.00	182.40	200.50	219.00
120	74.75	93.60	112.40	130.80	149.60	168.20	187.00	205.30	224.40

Length of tank (in.)

Width of tank (in.)	39	42	45	48	51	54	57	60	63
12	24.30	26.19	28.05	29.93	31.75	33.70	35.50	37.40	39.30
15	30.40	32.70	35.10	37.40	39.71	42.10	44.40	46.75	49.10
18	36.50	39.24	42.10	44.90	47.70	50.50	53.25	56.10	58.90
21	42.60	45.80	49.10	52.30	55.70	58.80	62.20	65.40	68.70
24	48.70	52.30	56.10	59.80	63.60	67.30	71.00	74.80	78.40
27	54.70	58.90	63.10	67.30	71.50	75.70	80.00	84.10	88.40
30	60.80	65.50	70.20	74.80	79.60	84.10	88.80	93.50	98.20
33	66.90	72.00	77.20	82.30	87.50	92.60	97.70	102.90	108.00
36	72.90	78.50	84.20	89.80	93.40	101.00	106.60	112.20	117.70
39	79.00	85.10	91.10	97.20	103.30	109.40	115.50	121.60	127.60
42	85.10	91.60	98.10	104.70	111.20	117.80	124.40	131.00	137.40
45	91.10	98.10	105.20	112.20	119.20	126.20	133.20	140.20	147.20
48	97.20	104.70	112.20	119.70	127.20	134.60	142.00	149.50	157.00
51	103.30	111.20	119.20	127.20	135.00	143.00	151.00	158.90	166.80
54	109.40	117.80	126.20	134.60	143.00	151.50	159.90	168.30	176.60
57	115.50	124.40	133.20	142.00	151.00	159.90	168.80	177.50	186.50
60	121.60	131.00	140.20	149.50	158.90	168.30	177.50	187.00	196.40
63	127.60	137.40	147.20	157.00	166.80	176.60	186.50	196.40	206.00
66	133.90	144.00	154.10	164.60	174.60	185.10	195.50	205.50	215.80
69	139.80	150.50	161.20	172.00	182.60	193.50	204.50	215.00	225.30
72	146.00	157.10	168.10	179.50	190.50	202.00	213.00	224.30	235.20
75	152.00	163.50	175.20	187.00	198.50	210.30	222.00	233.30	245.00
78	158.00	170.20	182.20	194.50	206.40	219.00	231.00	243.00	255.00
81	164.10	176.60	189.30	202.00	214.50	227.00	239.90	252.30	264.90
84	170.20	183.30	196.30	209.40	222.50	235.60	249.00	261.60	274.90
87	176.40	289.80	203.30	217.00	230.50	244.00	257.80	271.00	284.60
90	182.50	196.40	210.30	224.40	238.50	252.50	266.90	280.50	294.10
93	188.50	203.00	217.20	231.90	246.10	261.00	275.20	289.80	304.00
96	194.40	209.50	224.40	239.40	254.10	269.30	284.00	299.30	314.00
99	200.50	216.00	231.50	247.00	262.10	277.80	293.00	308.80	324.00
102	206.50	222.50	238.50	254.30	270.00	286.00	302.00	318.00	333.20
105	212.50	229.00	245.30	261.50	278.00	294.50	311.00	327.20	343.20
108	218.60	235.60	252.70	269.30	286.00	303.00	319.50	336.60	353.00
111	224.80	242.00	259.80	276.00	294.00	311.00	328.80	346.00	362.90
114	230.90	248.70	266.90	284.20	302.00	319.80	337.80	355.30	372.90
117	237.00	255.00	273.90	291.50	310.00	328.20	346.50	365.00	382.90
120	242.90	261.80	280.70	299.10	318.00	336.60	355.30	374.00	392.20

Length of tank (in.)

Width of tank (in.)	66	69	72	75	78	81	84	87	90
12	41.10	43.10	44.90	46.75	48.60	50.50	52.40	54.20	56.10
15	51.40	53.75	56.10	58.40	60.75	63.10	65.50	67.80	70.20
18	61.70	64.50	67.30	70.10	72.90	75.70	78.60	81.30	84.20
21	72.00	75.20	78.50	81.80	85.10	88.30	91.60	94.80	98.20
24	82.30	85.90	89.70	93.40	97.20	100.[illegible]	104.60	108.30	112.20
27	92.50	96.80	100.90	105.10	109.30	113.50	117.60	122.00	126.10
30	102.80	107.40	112.20	117.00	121.60	12[illegible].30	131.00	135.60	140.30
33	113.30	118.40	123.50	128.60	133.70	[illegible]	144.00	149.20	154.50
36	123.40	129.00	134.60	140.10	145.90	15[illegible].40	157.10	162.60	168.30
39	133.90	.139.80	146.00	152.00	158.00	[illegible].10	170.20	176.40	182.50
42	144.00	150.50	157.10	163.50	170.20	[illegible].60	183.30	189.80	196.40
45	154.10	161.20	168.10	175.20	182.20	[illegible].30	196.30	203.30	210.30
48	164.60	172.00	179.50	187.00	194.50	[illegible].00	209.40	217.00	224.40
51	174.60	182.60	190.50	198.50	206.40	[illegible]14.50	222.50	230.50	238.50
54	185.10	193.50	202.00	210.30	219.00	[illegible]7.00	235.60	244.00	252.50
57	195.50	204.50	213.00	222.00	231.00	[illegible].90	240.00	257.80	266.90

Length of tank (in.)

Width of tank (in.)	66	69	72	75	78	81	84	87	90
60	205.50	215.00	224.30	233.30	243.00	252.30	261.60	271.00	280.50
63	215.80	225.30	235.20	245.00	255.00	264.90	274.90	284.60	294.10
66	226.30	236.80	247.00	257.00	267.40	277.90	288.00	298.00	308.60
69	236.80	247.40	258.00	269.00	279.80	290.40	301.50	312.00	322.80
72	247.00	258.00	269.30	280.40	291.60	302.90	314.20	325.00	336.50
75	257.00	269.00	280.40	292.20	303.90	315.80	327.00	339.00	350.40
78	267.40	279.80	291.60	303.90	316.00	328.00	340.30	352.80	364.70
81	277.90	290.40	302.90	315.80	328.00	341.00	353.50	366.00	379.00
84	288.00	301.50	314.20	327.00	340.30	353.50	366.50	379.90	393.00
87	298.00	312.00	325.00	339.00	352.80	366.00	379.90	393.00	406.80
90	308.60	322.80	336.50	350.40	364.70	379.00	393.00	406.80	421.00
93	319.00	333.00	347.50	362.00	377.00	391.50	406.00	421.00	435.00
96	329.00	342.10	359.00	373.90	389.00	403.50	419.00	433.50	449.00
99	339.30	355.00	370.70	385.80	401.00	416.00	432.00	447.00	462.50
102	349.60	365.80	381.40	397.00	413.20	428.50	445.00	460.70	477.00
105	360.00	376.80	392.90	408.00	425.50	442.00	458.00	474.50	491.00
108	370.30	387.20	404.00	421.00	437.80	454.00	471.10	487.00	505.00
111	380.60	398.00	415.00	432.00	449.00	467.00	484.00	501.50	519.00
114	391.00	408.00	426.40	444.00	462.00	479.00	497.50	515.00	533.00
117	401.70	419.00	437.50	456.00	474.00	492.00	510.90	528.00	547.50
120	411.50	430.70	449.00	467.00	486.00	504.90	523.00	542.00	561.00

Length of tank (in.)

Width of tank (in.)	93	96	99	102	105	108	111	114	117
12	58.00	59.80	61.70	63.60	65.40	67.30	69.20	71.00	72.80
15	72.45	74.80	77.20	79.60	81.90	84.20	86.60	88.90	91.30
18	86.90	89.80	92.60	95.50	98.30	101.20	103.90	106.70	109.50
21	101.40	104.50	108.00	111.20	114.50	117.80	121.00	124.40	127.50
24	115.70	119.60	123.50	127.20	131.00	134.60	138.30	142.10	145.90
27	130.40	134.50	138.90	143.00	147.30	151.50	155.50	159.80	164.00
30	145.00	149.60	154.40	158.90	163.60	168.30	173.00	177.70	182.40
33	159.50	164.50	169.70	174.50	179.60	184.90	190.00	195.00	200.50
36	174.00	179.50	185.20	190.70	196.40	202.00	205.20	213.20	219.00
39	188.50	194.40	200.50	206.50	212.50	218.60	224.80	230.90	237.00
42	203.00	209.50	216.00	222.50	229.00	235.60	242.00	248.70	255.00
45	217.20	224.40	231.50	238.50	245.30	252.70	259.80	266.90	273.90
48	231.90	239.40	247.00	254.30	261.50	269.30	276.60	284.20	291.50
51	246.10	254.10	262.10	270.00	278.00	286.00	294.00	302.00	310.00
54	261.00	269.30	277.80	286.00	294.50	303.00	311.00	319.80	328.20
57	275.20	284.00	293.00	302.00	311.00	319.80	328.80	337.80	346.50
60	289.80	299.30	308.80	318.00	327.20	336.60	346.00	355.30	365.00
63	304.00	314.00	324.00	333.20	343.20	353.00	362.90	372.90	382.90
66	319.00	329.00	339.30	349.60	360.00	370.30	380.60	391.00	401.70
69	333.00	342.10	355.00	365.80	376.80	387.20	398.00	408.00	419.00
72	347.50	359.00	370.70	381.40	392.90	404.00	415.00	426.40	437.50
75	362.00	373.90	385.80	397.00	408.00	421.00	432.00	444.00	456.00
78	377.00	389.30	401.00	413.20	425.50	437.80	449.00	462.00	474.00
81	391.50	403.50	416.00	428.50	442.00	454.00	467.00	479.00	492.00
84	406.00	419.00	432.00	445.00	458.00	471.10	484.00	497.50	510.90
87	421.00	433.50	447.00	460.70	474.50	487.00	501.50	515.00	528.00
90	435.00	449.00	462.50	477.00	491.00	505.00	519.00	533.00	547.50
93	449.00	463.00	478.00	492.00	507.00	521.00	536.00	550.00	564.90
96	463.00	478.80	494.00	508.70	524.00	539.00	554.00	568.50	584.00
99	478.00	493.50	509.80	525.00	540.00	556.00	571.00	587.00	602.00
102	492.00	508.70	525.00	540.90	557.00	572.00	587.50	604.00	620.00
105	507.00	524.00	540.00	557.00	572.50	589.00	605.00	622.00	638.00
108	521.00	539.00	556.00	572.00	589.00	606.00	622.50	639.00	657.00
111	536.00	554.00	571.00	587.50	605.00	622.50	639.00	657.00	674.00
114	550.00	568.50	587.00	604.00	622.00	639.90	657.00	675.80	693.00
117	564.90	584.00	602.00	620.00	638.00	657.00	674.00	693.00	712.00
120	579.00	598.00	617.00	636.00	654.00	673.00	691.00	711.00	729.90

Width of tank (in.)	Length of tank (in.)						
	120	123	126	129	132	135	138
12	74.75	76.70	78.50	80.40	82.25	84.10	86.00
15	93.60	95.90	98.30	100.50	103.00	105.30	107.60
18	112.40	115.20	118.00	120.80	123.50	126.50	129.20
21	130.80	134.10	137.40	140.50	143.90	147.20	150.50
24	149.60	153.50	157.10	160.90	164.50	168.30	172.00
27	168.20	172.50	176.70	180.90	185.00	189.40	193.50
30	187.00	191.80	196.40	201.30	205.70	210.50	215.10
33	205.30	210.50	215.50	220.90	226.00	231.00	236.00
36	224.40	230.00	235.60	241.00	246.90	252.50	258.10
39	242.90	249.00	255.00	261.00	267.00	273.30	279.70
42	261.80	268.50	274.90	281.00	288.00	294.40	301.10
45	280.70	287.50	294.50	302.00	308.00	315.50	322.80
48	299.10	304.80	314.10	321.80	329.00	336.80	344.00
51	318.00	325.90	333.80	341.80	349.80	357.80	365.80
54	336.60	345.00	353.40	362.00	370.30	378.80	387.00
57	355.30	364.30	373.00	382.00	391.00	400.00	409.00
60	374.00	383.80	392.80	402.50	411.40	421.00	430.30
63	392.20	402.00	412.00	422.00	432.00	441.80	451.80
66	411.50	421.50	432.00	442.00	452.50	463.00	473.00
69	430.70	441.00	452.00	462.50	473.00	484.00	495.00
72	449.00	460.00	471.30	482.00	493.50	505.00	516.00
75	467.00	479.00	491.00	502.00	514.00	526.00	537.50
78	486.00	498.00	510.50	522.50	535.00	547.00	559.00
81	504.90	517.00	530.00	542.50	555.00	567.50	580.50
84	523.00	537.00	549.80	562.50	576.00	589.00	602.00
87	542.00	556.00	569.50	582.50	597.00	610.00	623.00
90	561.00	575.00	589.00	603.00	617.00	631.00	645.00
93	579.00	593.00	607.50	622.50	637.00	652.00	666.00
96	598.00	614.00	628.00	644.00	658.30	674.00	688.00
99	617.00	632.50	648.00	663.00	679.00	694.90	711.00
102	636.00	652.00	667.50	684.00	699.00	715.80	732.00
105	654.00	671.00	687.50	703.00	719.50	736.50	752.50
108	673.00	690.00	707.50	723.50	741.00	757.50	774.00
111	691.00	709.00	726.00	743.00	760.00	777.50	795.00
114	711.00	729.00	747.00	764.00	782.00	800.00	818.00
117	729.90	748.00	766.00	784.00	802.50	820.00	839.00
120	747.50	767.00	785.00	804.00	823.00	842.00	866.00

d. Elliptical Tanks (Horizontal).

(1) *Formula.* The full capacity in gallons of horizontal elliptical tanks with flat ends can be computed by the following formula.

$$\text{Volume (gal.)} = \frac{.7854abL}{231}$$

Where

a = long axis of elliptical cross-section

b = short axis of elliptical cross-section

L = length of tank

(2) *Table.* For tanks of known capacity, the partial content in gallons for varying liquid depths can be determined by using the scale below. (See *b*(2) above for example of use.)

Part of tank depth filled	Part of tank capacity filled	Part of tank depth filled	Part of tank capacity filled	Part of tank depth filled	Part of tank capacity filled
0.01	0.0020	0.05	0.0187	0.09	0.0445
0.02	0.0050	0.06	0.0245	0.10	0.0520
0.03	0.0090	0.07	0.0307	0.11	0.0598
0.04	0.0134	0.08	0.0374	0.12	0.0680

Part of tank depth filled	Part of tank capacity filled	Part of tank depth filled	Part of tank capacity filled	Part of tank depth filled	Part of tank capacity filled
0.13	0.0764	0.43	0.4114	0.73	0.7814
0.14	0.0850	0.44	0.4240	0.74	0.7927
0.15	0.0940	0.45	0.4366	0.75	0.8039
0.16	0.1032	0.46	0.4492	0.76	0.8150
0.17	0.1127	0.47	0.4619	0.77	0.8260
0.18	0.1224	0.48	0.4745	0.78	0.8368
0.19	0.1323	0.49	0.4873	0.79	0.8474
0.20	0.1423	0.50	0.5000	0.80	0.8577
0.21	0.1526	0.51	0.5127	0.81	0.8677
0.22	0.1632	0.52	0.5255	0.82	0.8776
0.23	0.1740	0.53	0.5381	0.83	0.8873
0.24	0.1850	0.54	0.5508	0.84	0.8968
0.25	0.1961	0.55	0.5634	0.85	0.9060
0.26	0.2073	0.56	0.5760	0.86	0.9150
0.27	0.2186	0.57	0.5886	0.87	0.9236
0.28	0.2300	0.58	0.6011	0.88	0.9320
0.29	0.2407	0.59	0.6136	0.89	0.9402
0.30	0.2531	0.60	0.6261	0.90	0.9480
0.31	0.2648	0.61	0.6386	0.91	0.9555
0.32	0.2766	0.62	0.6510	0.92	0.9626
0.33	0.2884	0.63	0.6634	0.93	0.9693
0.34	0.3003	0.64	0.6756	0.94	0.9755
0.35	0.3119	0.65	0.6881	0.95	0.9813
0.36	0.3244	0.66	0.6997	0.96	0.9866
0.37	0.3366	0.67	0.7116	0.97	0.9910
0.38	0.3490	0.68	0.7234	0.98	0.9950
0.39	0.3614	0.69	0.7352	0.99	0.9980
0.40	0.3739	0.70	0.7469	1.00	1.0000
0.41	0.3864	0.71	0.7593		
0.42	0.3989	0.72	0.7700		

e. Spherical Tanks. The chart ((4) below) may be used to determine the liquid volume in spherical tanks for various depths of the liquid. Also shown are three formulas which may be used instead of the table. When more precise computations are desired, use the formulas which do not require figures from the chart. Volumes are shown in cubic inches; cubic inches are divided by 231 to obtain gallons. In this subparagraph, the words "sphere" and "tank" have the same meaning, and the letters used represent the following:

d = depht of liquid or height of segment of sphere formed by liquid (when tank is less than half full)

= distance from top of sphere to surface of liquid (when tank is more than half full)

C = a value for different d/D relationships which represents the volume of the sphere segment divided by the cube of the sphere's diameter

D = diameter of tank

(1) *When tank is less than half full:*

(*a*) To find the volume of the liquid, form the ratio d/D and find the value of C in the table. Then Volume of liquid $= D^3C$

(*b*) Alternate method: (chart not required) Volume of liquid $= .5236d^2(3D-2d)$

(2) *When tank is half full:*

Volume of liquid $= .2618D^3$ (chart not required)

(3) *When tank is more than half full:*

(*a*) Since the ratio d/D (if d is considered the depth of liquid) becomes greater than 0.50, the table no longer applies. Therefore, the chart should be used to determine the volume of the unfilled portion. Subtract this from the total volume of the tank ($.5236D^3$) to obtain the volume of the filled portion. In this

case, d becomes the distance from the top of tank to the surface of liquid. To find the volume of the unfilled portion, form the ratio d/D and find the value of C in the chart. The unfilled portion is then D^3C, and the volume of the filled portion can be determined as follows:

$$\text{Volume of liquid} = .5236D^3 - D^3C = D^3\ (.5236 - C)$$

(*b*) Alternate method: (chart not required)

$$\left[\begin{array}{l}\text{Volume of liquid} = .5236D^3 \\ \quad -\ .5236d^3\ (3D - 2d)\end{array}\right]$$

(4) *Table.*

d/D	C	d/D	C	d/D	C	d/D	C	d/D	C
0.01	0.0002	0.11	0.0176	0.21	0.0596	0.31	0.1198	0.41	0.1919
0.02	0.0006	0.12	0.0208	0.22	0.0649	0.32	0.1265	0.42	0.1995
0.03	0.0014	0.13	0.0242	0.23	0.0704	0.33	0.1334	0.43	0.2072
0.04	0.0024	0.14	0.0279	0.24	0.0760	0.34	0.1404	0.44	0.2149
0.05	0.0038	0.15	0.0318	0.25	0.0818	0.35	0.1475	0.45	0.2227
0.06	0.0054	0.16	0.0359	0.26	0.0878	0.36	0.1547	0.46	0.2305
0.07	0.0073	0.17	0.0403	0.27	0.0939	0.37	0.1620	0.47	0.2383
0.08	0.0095	0.18	0.0448	0.28	0.1002	0.38	0.1694	0.48	0.2461
0.09	0.0120	0.19	0.0495	0.29	0.1066	0.39	0.1768	0.49	0.2539
0.10	0.0147	0.20	0.0545	0.30	0.1131	0.40	0.1843	0.50	0.2618

f. Temperature Corrections (Approximate) for POL Products.

Product	*Coefficient of expansion or contraction (base of 60° F.)**
Aviation gasoline	0.00070
Motor gasoline and naphtha (other than cleaning solvent)	0.00060
Light crude oil (above 35° API gravity), jet fuel, cleaning solvent, kerosene, distillate fuel oil, and fog oil	0.00050
Medium crude oil (15° to 35° API gravity, and lubricating oil	0.00040
Heavy crude oil (up to 15° API gravity), residual oil, and asphalt	0.00035

*Refers to change in unit volume per degree over or under 60° F.

Dimensions of Containers

Nomenclature	*Units in package*	*Type of package*	*Size of package (in.)* Length	Width or diameter	Height
Drum:					
U.S. 55-gal., 16-gage	1	Drum	0	24½	34¾
U.S. 55-gal., 18-gage	1	Drum	0	24⅝	34½
Can:					
U.S. 5-gal. (gasoline)	1	Can	13¾	6½	18⅝
U.S. 5-gal. (oil)	2	Case	0	11 15/16	14 3/16
U.S. 5-qt. (oil)	6	Case	0	14	10
U.S. 1-qt. (oil)	12	Case	18	13	6
Pail:					
U.S. 25-lb (grease)	1	Pail	0	11½	11½

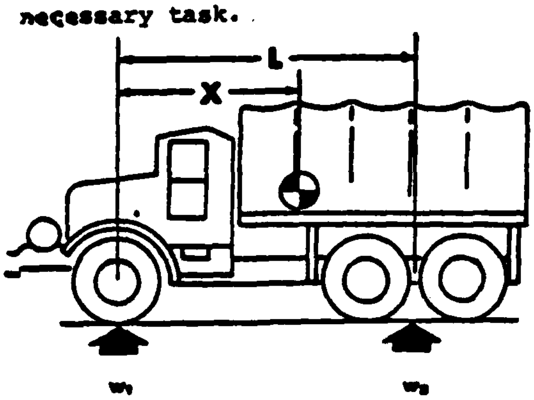
Bulk Capacities

Carrier	*Gallons*	*Short tons*	
		Gasoline 91A (bulk)	*Lube oil (bulk)*
Barge, coastwise [1]	200,000 to 400,000	45.9 to 91.8	761 to 1,522
Barge, harbor and canal [2]	15,000 to 30,000	257	57 to 114
Barge, Navy ponton [3]	84,000		320
Pipeline: [4]		930	
4-inch	304,000 per day	2,000	1,150
6-inch	655,000 per day [5]	3,500	2,500
8-inch	1,135,000 per day	614 to 1,228	4,350
Railroad tank car	8,000; 10,000; 12,000	24.1; 30.6; 36.8	30.4; 38.1; 45.7
Semitrailer, 12-ton, 4W	5,000	15.3	19
Ship, large tanker [6]	2.5 to 11 million	7,620 to 33,500	9,480 to 43,800
Ship, small tanker [7]	600,000 to 2 million	1,830 to 6,140	2,280 to 7,610
Tank, bolted-steel	10,500; 42,000; 420,000	32.2; 128; 1,280	39.9; 160; 1,600
Tank, portable fabric [8]	*10,000*	*30.6*	38.1
Tank truck, F-3, fuel or oil	750	2.3	2.9
Tank truck, L-2, oil service	600	1.8	2.3
Trailer, fuel-servicing	600	1.8	2.3
Transporter, liquid, rolling-wheel type (RLT), 1,000-gal T3 [9]	1,000	3.0	3.7
Truck, tractor and trailer, F-1	4,000	12.2	15.2
Truck, tractor and 2 trailers, F-1A	8,000	24.4	30.4
Truck, tractor and trailer, F-2	2,000	6.1	7.6
Truck, tractor and 2 trailers, F-2A	4,000	12.2	15.2

[1] Molded hulls.
[2] Rectangular hulls.
[3] 6x18 ponton barge carrying three 42,000-gallon tanks loaded to two-thirds capacity.
[4] In maintaining the same volumetric pipeline capacity for gasoline and oil, more pressure is required for the heavier liquid.
[5] Based on 32,500 gallons per hour for 20 hours of operation. In an emergency it can deliver 30,000 gallons per hour for 24 hours of operation, or 720,000 gallons per day.
[6] The ship tanker most commonly used is the T2-SE-A1, a 5,922,000-gallon tanker. It is 425 feet long and draws 31 feet. It has three 8-flanged discharge outlets and four discharge pumps rated 1,000 gpm at 100 psi.
[7] Draft loaded, 12 to 20 feet.
[8] 40 feet long, 12 feet wide, 3 feet high when filled. When empty, it can be rolled to 20 inches by 12 feet; 10 can be carried in a 6x6 truck.
[9] A pair of removable synthetic-rubber containers (fuel cells) mounted on an axle and towing unit. Each cell has a capacity of 500 gallons.

5-6 CENTER-OF-GRAVITY DETERMINATION

The military transportation officer is often confronted with the problem of shipping vehicles or other wheeled devices. This almost always involves the computation of the vehicle's center-of-gravity. The following descriptions provide a simple step-by-step approach to accomplish this very necessary task.

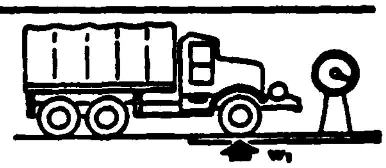


$$X = \frac{L\,w_2}{W}$$

WHERE

- X = Distance from front axle to unit center-of-gravity location
- L = Wheelbase
- w_1 = Front axle load
- w_2 = Rear axle load
- W = Total weight of unit

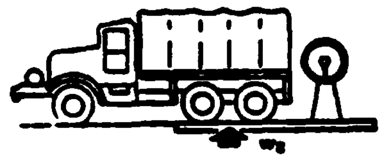
TO LOCATE CARGO UNIT CENTER OF GRAVITY, DETERMINE X



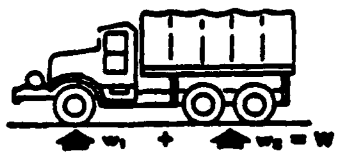
1. Determine axle loads by weighing all axles (w_1 and w_2).

   Note

   Vehicle must be level when weighing.

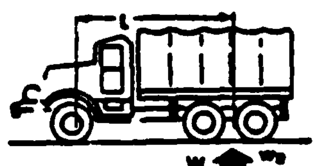


2. Determine total weight of unit (W) by adding axle loads (w_1 and w_2).

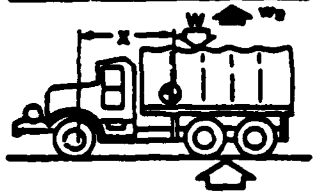


3. Determine wheelbase (L).

4. Determine rear axle moment about front axle by multiplying rear axle load (w_2) by wheelbase (L).



5. Determine center-of-gravity distance from front axle (X) by dividing rear axle moment by total weight (W).

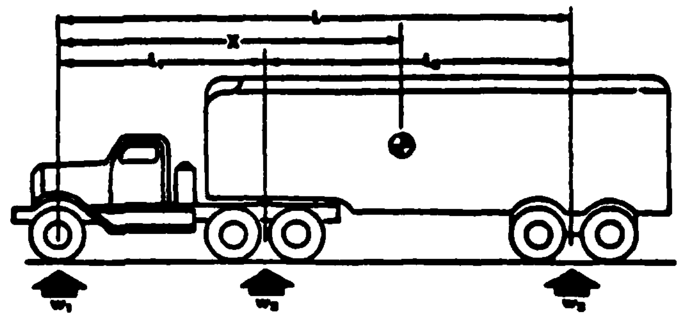


REAR AXLE MOMENT

$$X = \frac{L_1 w_2 + L\ w_3}{W}$$

WHERE

X = Distance from front axle to unit center-of-gravity location
w_1 = Front axle load
L_1 = Tractor wheelbase
w_2 = Tractor rear axle load
L_2 = Trailer wheelbase
w_3 = Trailer axle load
L = Total wheelbase of unit
W = Total weight of unit



TO LOCATE CARGO UNIT CENTER OF GRAVITY, DETERMINE X

---

1. Determine axle loads by weighing all axles (w_1, w_2, and w_3).

Note

Vehicle must be level when weighing

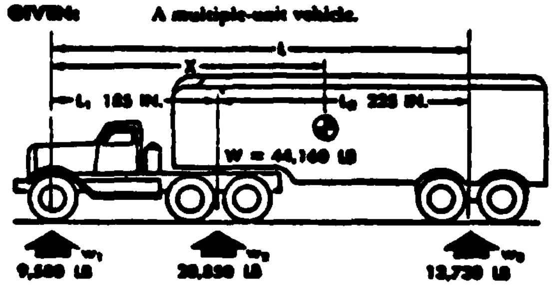
2. Determine total weight of unit (W) by adding all axle loads.
3. Determine tractor wheelbase (L_1) and trailer wheelbase (L_2).
4. Determine total wheelbase of unit (L).
5. Determine tractor rear axle moment about front axle by multiplying tractor wheelbase (L_1) by tractor rear axle load (w_2).
6. Determine trailer axle moment by multiplying total wheelbase (L) by trailer axle load (w_3).
7. Determine total moment about front axle by adding trailer axle moment and tractor rear axle moment.
8. Determine center-of-gravity distance from front axle (X) by dividing total moment by total weight (W).

Note:

As an aid to load planning, it may be desirable to know relationship of vehicle center of gravity to vehicle extremities. To determine distance from front bumper to center of gravity, add distance between front bumper and front axle to value determined in step 8. To determine distance between aft end of vehicle and center of gravity, subtract value determined in step 8 from distance between front axle and aft end of vehicle.

SAMPLE PROBLEM

GIVEN: A multiple-unit vehicle.



PROBLEM: Determine center-of-gravity location.

SOLUTION:

1. w_1 = 9,580 pounds
   w_2 = 20,850 pounds
   w_3 = 13,730 pounds
2. W = 9,580 + 20,850 + 13,730 = 44,160 pounds
3. L_1 = 185 inches
   L_2 = 225 inches
4. L = 185 + 225 = 410 inches
5. 185 × 20,850 = 3,857,250
6. 410 × 13,730 = 5,629,300
7. 3,857,250 + 5,629,300 = 9,486,550
8. 9,486,550 ÷ 44,160 = 214.8

CONCLUSION:

Cargo unit center of gravity is located 214.8 inches aft of front axle.

MULTIPLE—UNIT VEHICLES

5-7 MISCELLANEOUS DATA

The following lists contain information that is particularly useful in the field of military transportation.

DIMENSIONS OF CONTAINERS

Nomenclature	Units in Package	Type of Package	Length	Size of package (in.) Width or Diameter	Height
Drum:					
U.S. 55-gal., 16-gage	1	Drum	0	24 1/8	34 3/4
U.S. 55-gal., 18 gage	1	Drum	0	24 5/8	34 7/8
Can:					
U.S. 5-gal., (gasoline)	1	Can	13 3/4	6 1/2	18 3/8
U.S. 5-gal., (oil)	2	Case	0	11 15/16	14 3/16
U.S. 5-qt. (oil)	6	Case	0	14	10
U.S. 1-qt., (oil)	12	Case	18	13	6
Pail:					
U.S. 25-lb (grease)	1	Pail	0	11 1/2	11 1/2

WEIGHTS OF LIQUIDS AND GASES

Fluid in Pipeline	Pounds per cu ft
Liquids:	
Aviation gasoline	44.1
91A gasoline	45.7
Kerosene	50.8
Diesel oil	52.3
Lubricating oil	56.8
Water, fresh	62.4
Water, sea	64.0
Gases:	
Air	0.07658
Ammonia	0.04509
Benzene	0.20640
Butane	0.15350
Carbon dioxide	0.11637
Chlorine	0.18750
Ethylene	0.07410
Helium	0.01058
Hydrogen	0.00530
Methane	0.04234
Natural gas	0.05140
Oxygen	0.08463
Propane	0.11645
Steam	0.04761

CHAPTER 6

BASIC STATISTICAL CONCEPTS

INTRODUCTION

Basic statistical concepts which are of use to the military logistician are reviewed in this chapter. Included under each topic area are a general description of the technique, algebraic principles and formulas, and at least one example of a specific application.

The techniques presented in this chapter should not be considered as the only input to the decision-making process. Statistics can neither prove nor disprove anything. Their primary purpose is to provide a clear, logical approach to: (1) the conversion of raw data into meaningful information, (2) provide support or credibility to a particular management decision, or (3) narrow the list of alternative solutions to a problem.

DEFINITIONS

Some of the basic terms used in the discussion of probability and statistics are defined below.

Experiment. An experiment is any process of observation. A random experiment is one in which the outcome cannot be predicted with certainty, i.e., the outcome is a chance occurrence as in tossing a die.

Sample Space. The set of all possible outcomes of a random experiment. If the experiment is the toss of a die, the sample space, $S = \{1,2,3,4,5,6\}$.

Event. An event is a subset of the sample space and may include one or more of the possible outcomes. If an event contains exactly one sample point or outcome, it is called a simple event. If it contains more than one sample point, it is called a compound event. In a toss of a die, a compound event E_1, could be defined as an odd number. If a 1, 3, or 5 occurs, then event E_1 will have occurred.

Events are said to be mutually exclusive if they cannot occur simultaneously, i.e., if their intersection is the null or empty set ($\emptyset$).

Events are collectively exhaustive if the union of two or more events comprises the entire sample space.

When two or more events are both mutually exclusive and collectively exhaustive, they form a partition in which one of the set of outcomes of an experiment must occur, but only one can occur in any single trial.

Probability. Probability (P) is a real number which describes the likelihood of occurrence of an event (E) from a sample space (S) such that the following axioms hold:

1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. $P(E_i \cup E_j) = P(E_i) + P(E_j)$ where $P(E_i \cap E_j) = \emptyset$

A Priori. The classical definition of probability is based upon the notion of equally likely outcomes of an experiment. Probabilities are found through prior knowledge by enumerating the entire finite sample space. For example, in the toss of fair coin, it is known a priori that the probabiility of a head occurring is 1/2.

Relative Frequency. Probabilities can also be calculated by observing outcomes over a large number of trials when the probabilities are not known a priori. Thus, the relative frequency probability of an event E is defined as:

$$P(E) = \frac{m}{n}$$

where: m = the number of occurrences of event E
n = the total number of trials

Subjective. Probabilities may be determined judgmentally or based upon the personal experience of the manager of management group. For example, in order to assess the impact of pending teamster contract negotiations, a transportation officer could assign probabilities to the possible outcomes, such as:

Event	P
Settlement	.6
Lengthy arbitration	.3
Labor strike	.1
	1.0

Note that the foregoing example constitutes a probability distribution for the possible outcomes in that:

1. The probabilities sum to unity, i.e., $\Sigma P = 1$.
2. The events are mutually exclusive, i.e., no two events can occur at the same time.
3. The probability of each event is between zero and one.
4. The events are collectively exhaustive, i.e., based on the manager's experience; no other outcomes are possible in this case.

Representation of Outcomes. Probabilistic outcomes can be represented in standard set notation, tree diagrams, or Venn diagrams as follows:

Set — Tree Diagram — Venn Diagram

E.g. Single toss of coin:

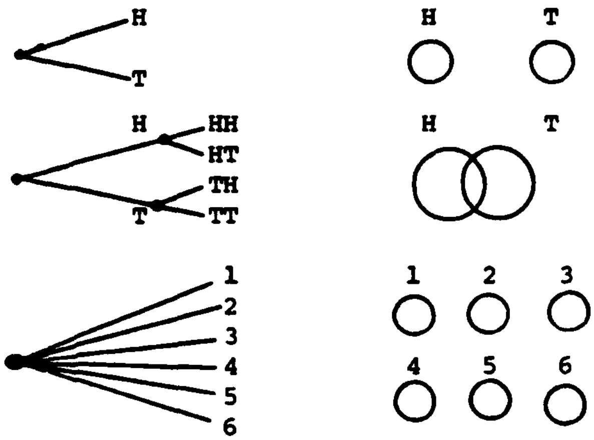
$S = \{H,T\}$

E.g. Toss coin twice:

$S = \{HH,HT,TH,TT\}$

E.g. Roll a die:

$S = \{1,2,3,4,5,6\}$



Counting Principles. Fundamentally, probabilities are computed by finding the ratio of favorable outcomes to total possible outcomes,

$$\text{or} \quad P(E) = \frac{m}{n}$$

which is the relative frequency definition provided earlier. To evaluate m and n, basic counting principles are used. Simply defined: If there are X different ways to accomplish event A, and Y different ways to to accomplish event B, then there are:

$X + Y$ ways to accomplish A or B

$X \cdot Y$ ways to accomplish A and B

Permutation. A set of elements arranged in some order. Consider, for example, the first four digits of a federal stock number. The number 1100 is different than 1010 which contains the same digits. Thus, when the order in which the values appear is important, the rules for permutations should be followed. In general, the permutation of n things taken r at a time is found by:

$$_nP_r = \frac{n!}{(n-r)!}$$

where $n!$ (n factorial) is the product of all the natural numbers from 1 to n. Thus, $n! = n(n-1)!$ and, by definition $0! = 1$. In the example of the first four digits of a stock number, assuming that digits cannot be repeated and that zero can occupy any position including the first, the number of possible stock numbers can be found by:

$$_{10}P_4 = \frac{10!}{(10-4)!} = \frac{10!}{6!} = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$$

In the special case when $n = r$, $_nP_r = n!$

Repetitions. Another special case occurs when permuted objects involve repetitions. For example, suppose a code requires the use of only the letters KBBRB. The number of 5-letter "words" which can be formed using those letters is found by:

$$_nP_{n_1 n_2 \ldots n_r} = \frac{n!}{n_1! n_2! \ldots n_r!}$$

which is the permutation of n objects of which n_1 are alike, n_2 are alike, ..., n_r are alike. In the example, the letter B appears 3 times, so there is no way to distinguish between them. The number of different words is:

$$\frac{5!}{3!1!1!} = 20$$

(no. of Bs) (no. of Ks) (no. of Rs)

Permutations are especially useful in statistical sampling.

Sampling with Replacement. How many ways can 5 cards be selected from a deck of 52 if each card is replaced after it is drawn? The total number of possible outcomes (sample space) is:

$$n^r = 52^5 = 380,204,032$$

Sampling without Replacement. If the cards are not replaced:

$$_nP_r = {}_{52}P_5 = \frac{52!}{(52-5)!} = 311,875,200$$

Combination: A set of n objects taken r at a time in which order is not important can be found by:

$$_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Note that $_nC_r = \left({}_nP_r\right)/r!$

EXAMPLE 6-1. A new USAF briefing team is being established which will consist of 2 majors and 2 captains. If there are 5 majors and 8 captains to choose from, in how many ways can the team be selected? Since order is not important, there are $_5C_2$ ways to choose majors and $_8C_2$ ways to choose captains or $_5C_2 \cdot {}_8C_2$ ways to select the team.

$$_5C_2 = \binom{5}{2} = \frac{5!}{3!2!} = 10$$

$$_8C_2 = \binom{8}{2} = \frac{8!}{6!2!} = 28$$

or 10.28 = 280 ways in which a team can be selected.

Computing Probabilities. Events are said to be independent when the outcome of one trial in no way influences the outcome of another. For example, in the toss of a fair coin, a head appearing on the first toss will not affect the probability of obtaining a head on the second toss. For independent events, the probability of the joint occurrence of two or more events is obtained via the special rule of multiplication. If A and B are independent events, then the probability of their joint occurrence (their intersection or $\cap$) is given by the product of their marginal probabilities. $P(A \cap B) = P(A) \cdot P(B)$. $P(A \cap B)$ is read, "the probability of both A and B occurring."

EXAMPLE 6-2. In an experiment consisting of two tosses of a coin, what is the probability of obtaining two heads? A priori it is known that the probability of obtaining a head (or tail) in a toss of a fair coin is 0.5. To obtain two heads in two tosses, appropriately assuming that each toss is independent,

$$
\begin{aligned}
P(H_1 \cap H_2) &= P(H_1) \cdot P(H_2) \\
&= 0.5 \cdot 0.5 \\
&= 0.25
\end{aligned}
$$

Conditional Probability is the probability that some event occurs given that some other event has already occurred, or $P(A|B)$ is read, "the probability of event A given that event B has already occurred." The conditional probability of A given B is found using the product rule:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{for } P(B) \neq 0$$

If A and B are independent,

$$P(A|B) = P(A)$$

and

$$
\begin{aligned}
P(A \cap B) &= P(A|B) \cdot P(B) \\
&= P(A) \cdot P(B) \text{ as above}
\end{aligned}
$$

An important aspect of conditional probability is reduced sample space. This occurs when samples are selected without replacement. It should also be noted that independent events are not mutually exclusive.

Other rules and relationships may be useful in determining probabilities.

Complements. $\bar{A}$ is read "not A". $P(\bar{A}) = 1 - (P(A)$

De Morgan's laws: $P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B}) = 1 - P(A \cap B)$

$$P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$$

EXAMPLE 6-3. Sixty-seven officers are available for reassignment as logisticians. Of these, 47 have had supply experience, 35 have had maintenance experience, and 23 have had both supply and maintenance experience. If the 66XX career counselor selects one at random, what is the probability that the officer selected has had neither supply nor maintenance experience?

Let S = event, officer had supply experience
M = event, officer had maintenance experience

It is known from above that

$$P(S) = \frac{47}{67} \qquad P(M) = \frac{35}{67} \qquad P(S \cap M) = \frac{23}{67}$$

We wish to find $P(\bar{S} \cap \bar{M})$ Using De Morgan's law:

$$P(S \cup M) = 1 - P(\bar{S} \cap \bar{M})$$

or $$P(\bar{S} \cap \bar{M}) = 1 - P(S \cup M)$$

Since $$P(S \cup M) = P(S) + P(M) - P(S \cap M)$$

Then $$P(\bar{S} \cap \bar{M}) = 1 - [P(S) + P(M) - P(S \cap M)]$$

$$= 1 - \frac{47 + 35 - 23}{67}$$

$$= 1 - .88$$

$$= 0.12$$

It can also be concluded that (.12)(67) or 8 of the 67 officers have had neither supply nor maintenance experience.

Bayes' Theorem. An extension of conditional probability states that ... "Given B_1, B_2 ... B_r mutually exclusive, collectively exhaustive events let A be an arbitrary event in the universe such that $P(A) \neq 0$, then

$$P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{\sum_{i=1}^{n} P(B_i) \cdot P(A|B_i)}$$

where $i = 1, 2, \ldots, n$
$j = 1, 2, \ldots,$ or n

determines the posterior probability, $P(B_j|A)$. Uncertainty is reduced through the use of additional information from past samples, experiments, etc., to modify prior probabilities.

EXAMPLE 6-4. A certain component is produced by three manufacturers. The total output is purchased by the Defense Logistics Agency. Of the total production, 45%, 20% and 35% are produced by manufacturers A, B, and C, respectively. From historical records, it is known that 10%, 5%, and 1% of the production is defective from manufacturers A, B, and C, respectively. If the user finds a part which is defective, what is the probability that it came from manufacturer A?

Let D = event, part is defective
A = event, part was made by A
B = event, part was made by B
C = event, part was made by C

What is $P(A|D)$?

Using Bayes' theorem

$$P(A|D) = \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|B)P(B) + P(D|C)\ P(C)}$$

$$= \frac{(.10)\ (.45)}{(.10)\ (.45) + (.05)\ (.2) + (.01)\ (.35)}$$

$$= 0.77$$

Using the information that "the part is defective," the prior probability that the part came from manufacturer A is revised from 0.45 to 0.77.

CHAPTER 7

PROBABILITY DISTRIBUTIONS

DISCRETE PROBABILITY DISTRIBUTIONS

Random variables can be described as outcomes of a probabilistic process that can be measured as a subset of the real numbers. Random variables having only integer values are known as discrete random variables.

The probability distribution of a discrete random variable k is described by its probability mass function P(k) which gives the probability of occurrence of each value of the random variable and must meet the following conditions:

$$o \leq P(k) \leq 1$$

$$P(k_i \cup k_j) = P(k_i) + P(K_j)$$

$$\sum P(k) = 1$$

A cumulative distribution function F(k) gives the probability that k will take on a value which is less than or equal to a specified value k_o within its domain.

$$F(k_o) = \sum_{k=a}^{k_o} P(k) = P(k \leq k_o)$$

where $a \leq k \leq b$

The mean and variance are often-used descriptors of probability distributions. The mean describes the central tendency of the data values in the distribution, i.e., where do most of the observations occur? The variance describes the spread of data values, or how far away from the mean are they located? In other words, the mean specifies the location of the distribution (on the number line) and the variance specifies the shape, i.e., flat or peaked.

Mean.

$$\mu_k = E(k) = \sum_k kP(k)$$

where μ_k = mean or average of all the values of k, a discrete random variable

$E(k)$ = expected value of k, another name for the mean

$P(k)$ = probability associated with each value of k.

Note that the mean is a weighted average in which the values of k are weighted by their respective probabilities.

Variance.

$$\sigma_k^2 = \mathrm{Var}(k) = E\left[(k-\mu_k)^2\right]$$

$$= \sum_k (k-\mu_k)^2 P(k)$$

The variance is referred to as the second moment and expresses the expected value of the squared deviations of k from its mean, also weighted by the probabilities of k.

A number of discrete probability distributions have been found to occur quite frequently in nature to describe the outcomes of random processes. Several of these which have specific application in the many functions of logistics are described below. The reader should be cautioned, however, that the theoretical probability distributions given herein may not completely describe a logistics process. In such cases, erroneous conclusions could be drawn from their use. In many instances, however, use of the following distributions can and have provided valuable inputs to the complex area of logistics decision making.

Uniform Distribution

Definition. A discrete probability distribution where all the integers from α through β have an equal probability of occurrence. This distribution stems from equally likely probability in which all values of a random variable are equally likely to occur.

Probability mass function.

$$P(k) = \frac{1}{\beta-\alpha+1}; \; k = \alpha, \alpha+1, \alpha+2, \cdots, \beta$$

where k is a discrete random variable

Parameters. α = the smallest value of k

β = the largest value of k

Mean: $$\mu_k = E(k) = \sum_{k=\alpha}^{\beta} \frac{k}{\beta-\alpha+1}$$

$$= \frac{\alpha+\beta}{2}$$

Variance. $$\sigma_k^2 = V(k) = \sum_{k=\alpha}^{\beta} \frac{k-\mu_k{}^{2}}{\beta-\alpha+1}$$

$$= \frac{(\beta-\alpha)(\beta-\alpha+2)}{12}$$

Cumulative distribution function.

$$F(k_o) = P(k \leq k_o) = (k_o-\alpha+1) \quad \frac{1}{\beta-\alpha+1}$$

where k_o is a particular specified value of the random variable k.

Uses.

1. This distribution could be used in a random sampling process as a means to determine whether or not a given sample is random.

2. In the absence of probabilistic information for any discrete process, the uniform distribution could be used to provide probabilities of the outcomes of the process.

EXAMPLE 7-1. At an aerial port the number of pallets of cargo awaiting airlift on any given day has an unknown distribution with no less than four nor more than eleven on hand.

a. What is the expected number of pallets on any given day?

$$\mu_k = E(k) = \frac{11+4}{2} = 7.5 \text{ pallets}$$

b. What is the probability that more than seven pallets will be awaiting airlift?

$$P(k > 7) = P(k \geq 8) = 1-P(k \leq 7)$$

$$= 1- (7-4+1) \frac{1}{11-4+1}$$

$$= 0.5$$

(Incidently, this is the same probability of having seven or less pallets awaiting airlift.)

This information might be useful to the traffic manager for scheduling airlift, loading equipment, and manpower during periods of normal operations. During emergency operations such as combat support, deployment, or relief operations as a result of a national disaster, this distribution would not be applicable as it assumes an even flow of pallets and could not account for peaks and valleys in the work load.

Binomial Distribution

Definition. A discrete probability distribution which expresses the probability of success associated with the outcomes of n Bernoulli trials, where:

1. A Bernoulli trial is an experiment with only two outcomes, usually termed "success" or "failure."

2. Each trial is independent.

3. T' ere is a constant probability of success for each trial.

4. There is a fixed number of repeated trials.

Probability mass function.

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where k = the number of successes, and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Parameters. n = the number of trials

p = the probability of success for each trial

Mean. $\mu_k = E(k) = np$

Variance. $\sigma_k^2 = V(k) = np(1-p)$

Cumulative distribution function. The probability of obtaining k_o or fewer successes in n trials can be found by:

$$F(k_o) = \sum_{0}^{k_o} \binom{n}{k} p^k (1-p)^{n-k}\ , \text{ for } k = 0,1,2,\cdots, k_o$$

Cumulative probabilities can be found more easily in widely published cumulative binomial probability tables.

Uses. This distribution may be used with any Bernoulli process (two outcomes) when sampling from an infinite (or very large) population, or when sampling with replacement from a finite population.

EXAMPLE 7-2. A contractor has produced a certain part in large quantities for the Defense Logistics Agency. Historically it has been found that 5 percent of the parts are defective. In the latest shipment of 30 parts, what is the probability that at least one defective will be found?

The probability of at least one defective is given by:

$P(k \leq 1)$ or

$1-P(k=0)$ where k is the number of defectives

$$= 1 - \left[\binom{30}{0} (0.05)^0\ (0.95)^{30}\right]$$

$= 0.786$

Poisson Distribution

Definition. This distribution is the limiting case of the binomial distribution and can be used to approximate the binomial particularly when $n > 100$ and $p < 0.10$. The Poisson is also the distribution of the number of independent occurrences of an event over a specified interval of time, volume, or space.

Probability mass function.

$$P(k) = \frac{(np)^k}{k!} e^{-np}$$, which is the limit of the binomial distribution as $n \to \infty$ and $p \to 0$.

where k = the number of occurrences of the event in the interval

n = total number of intervals in the sample

p = probability of an occurrence in the interval

More appropriately, the mass function is given by:

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

where $\lambda = np$

Mean. $\mu_K = E(k) = \lambda$

Variance. $V(k) = \sigma_k^2 = \lambda$

If the Poisson is used in instances other than for approximating the binomial, the following conditions must be met:

1. Events can occur at any point in the continuum of space or time.

2. There is a small probability of occurrence of a single event in a given interval.

3. The probability of occurrence is constant for each interval.

4. The number of occurrences in one interval is independent of the number in any other interval.

Uses. The Poisson distribution is often used to describe

1. the number of arrivals in a given time period at a queue;

2. the number of failures of a component during a given time period; and

3. the number of defects in a given length of material.

EXAMPLE 7-3. An airbase has an average of two unscheduled landings per hour for any given 24-hour period. If the runway must be closed for six hours for repair, what is the probability that two or less unscheduled landings must be diverted to another facility?

If an average of 2 unscheduled landings occur per hour then

$$\lambda = 2 \cdot 6 = 12$$

can be expected to occur in any 6-hour period. Letting $\lambda = 12$

$$P(k \le 2) = P(k = 0) + P(k = 1) + P(k = 2)$$

$$\text{or} \quad = \sum_{k=0}^{2} P(k)$$

$$= \frac{12^0}{0!} e^{-12} + \frac{12^1}{1!} e^{-12} + \frac{12^2}{2!} e^{-12}$$

$$= 0.0005$$

Conversely, there is a (1.0000 - .0005) X 100 or 99.95 percent chance that more than 2 unscheduled aircraft will arrive.

Hypergeometric Distribution

Definition. A discrete probability distribution which is somewhat like the binomial except that the items in a sample are drawn without replacement. The probability of success is not constant for each trial.

Probability mass function.

$$P(k) = \frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}} \; ; \; k = 0, 1, 2, \cdots, n$$

where k is a discrete random variable, the number of success sample.

Cumulative distribution function.

$$F(k_o) = P(k \le k_o) = \sum_{k=0}^{k_o} P(k)$$

$$\text{and} \quad \binom{N_1}{k} = \frac{N_1!}{k!(N_1-k)!}$$

$$\binom{N-N_1}{n-k} = \frac{(N-N_1)!}{(n-k)!(N-N_1-n+k)!}$$

Parameters.

N_1 = total number of successes in the population

n = sample size

N = population size

Mean. $\mu_k = E(k) = \dfrac{nN_1}{N}$

Variance. $\sigma_k^2 = V(k) = \left(\dfrac{nN_1}{N}\right)\left(1 - \dfrac{N_1}{N}\right)\left(\dfrac{N-n}{N-1}\right)$

Uses. The hypergeometric distribution is used when only two outcomes are possible for any given trial and when samples are drawn without replacement. When samples are drawn with replacement, the binomial distribution is used instead.

EXAMPLE 7-4. A squadron is preparing for a deployment exercise. A WRSK kit contains 10 of a particular replacement component. In past deployments, an average of 4 of the components has been found to be defective. Due to weight limitations, only 5 of the components will accompany the next deployment exercise. If 2 of the components will be used during the deployment, what is the probability that at least 2 of the 5 components in the kit will be serviceable? If k = the number of serviceable components in the "sample" of 5 taken on the deployement, the probability that at least 2 of the components are serviceable can be found by:

$$P(k \geq 2) \text{ or } 1 - P(k \leq 1)$$

$$= 1 - [P(k=o) + P(k=1)]$$

$$= 1 - \frac{\binom{6}{0}\binom{4}{5}}{\binom{10}{5}} + \frac{\binom{6}{1}\binom{4}{4}}{\binom{10}{5}}$$

$$= 1 - 0 + \frac{\left(\frac{6!}{1!5!}\right)\left(\frac{4!}{0!4!}\right)}{\left(\frac{10!}{5!5!}\right)}$$

$$= 1 - \frac{6}{252}$$

$$= 0.97619$$

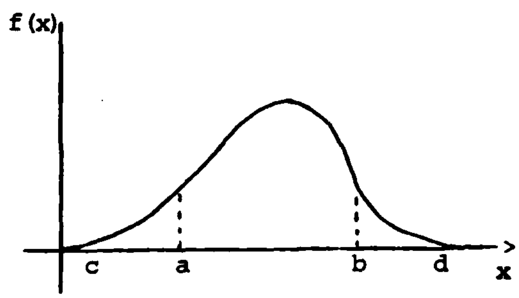
Continuous Probability Distributions

A continuous random variable can take on an infinite number of values over a sample space. For example, the number of weights between 180 and 181 pounds is theoretically infinite, but in a practical sense limited by the sensitivity of the weighing (or measurement) device.

Since a continuous random variable can take on an infinite number of values, the probability that a single value will be observed is essentially zero, or

$$P(X = X_o) = \frac{1}{\infty} \approx 0$$

The probability distribution of a continuous random variable is represented by a probability density function (PDF) called f(x). Graphically, the PDF for a continuous random variable x is illustrated below:



For a continuous random variable x, the PDF must exhibit the following properties

1. $f(x) \geq o$ for all x
2. $\int_c^d f\ (x)\ dx = 1$ $c \leq x \leq d$

Probability is measured by area under the curve described by f(x), known as the cumulative distribution function (CDF) or F(x), where

$$F(a) = P(x \leq a) = \int_c^a f(x)dx$$

which is the probability that the random variable is less than or equal to some value a. In addition to the properties of the PDF above, the following properties hold for the CDF.

1. $P(x < c) = F(c) = 0$

2. $P(x > d) = 0$

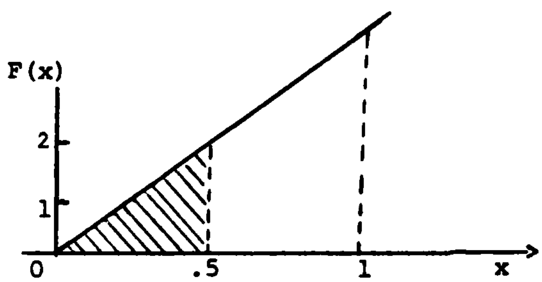
3. $P(x < d) = F(d) = 1$

It should be noted that the probability density function is the first derivative of the cumulative distribution function.

$$f(x) = \frac{dF(x)}{dx}$$

Consider the following CDF:

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$



The shaded area represents the probability that a continuous random variable x is less than 0.5, and can be found by:

$$F(0.5) = P(x < 0.5) = \int_0^{0.5} f(x)\ dx$$

$$= \int_0^{0.5} 2x dx$$

$$= \quad F(0.5) - F(0) = 0.25$$

Note that $2x = \frac{d[x^2]}{dx}$ or $f(x) = \frac{dF(x)}{dx}$

The mean and variance are useful descriptors of any probability distribution. These can be found using the methodology described above.

Mean. $\mu_x = E(x) = \int_c^d x f(x)\,dx$

Using the example above:

$$\mu_x = \int_o^1 (x)(2x)\,dx$$

$$= \int_o^1 2x^2 dx$$

$$= \left[\frac{2}{3}\ x^3\right]_o^1$$

$$= F(1) - F(0\ )$$

$$= \frac{2}{3} - 0$$

$$= 0.667$$

Variance.

$$\sigma_x^2 = \int_c^d (x-\mu_x)^2\ f(x)\,dx$$

$$= \int_c^d x^2 f(x)\,dx - \mu_x{}^2$$

Using the example above:

$$\sigma_x^2 = \int_o^1 (x^2)(2x)\,dx - \mu_x{}^2$$

$$= \int_o^1 2x^3 dx - \mu_x{}^2$$

$$= \left[\frac{x^4}{2}\right]_o^1 - \left(\frac{2}{3}\right)^2$$

$$= F(1) - F(0) - \frac{4}{9}$$

$$= \tfrac{1}{2} - 0 - \frac{4}{9}$$

$$= \frac{1}{18}$$

$$= 0.056$$

The probability distributions of continuous random variables and hence their probabilities can be found in a similar manner. However, there are a number of probability distributions which occur quite often in the area of logistics as well as others. Probabilities have already been computed and can be found in published tables. The more common continuous probability distributions useful to logisticians are contained in the sections which follow.

Normal Distribution

Definition. A continuous probability distribution described by a smooth, symmetric, bell-shaped curve which is completely specified by its mean and variance. One of the most useful probability models, it is characterized by the following:

1. It is symmetric about its mean, μ.

2. It has an infinite range, i.e., unbounded, so that any interval of numbers will have a positive probability.

3. The density curve falls off quickly as values deviate from the mean. In fact, more than 99 percent of the area under the normal curve is encompassed by $\mu \pm 3\sigma$, which makes this distribution a good approximation of other distributions such as the binomial or Poisson.

4. The distribution is completely defined by its mean and variance. A change in the mean displaces the distribution to the right or left; a change in the variance alters its shape.

5. It exhibits the property of additivity; i.e., the sum of independent normal variables is also a normal variable.

6. Probabilities are computed from the standard normal distribution which has a mean of zero and variance of one and can be defined as:

$$Z = \frac{X-\mu}{\sigma}$$

where

Z = standard normal variable

X = a normally distributed variable

μ = mean of the normally distributed variable

σ = standard deviation of the normal variable

Probability density function (standard normal).

$$f(Z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}Z^2} \quad \text{for } -\infty < Z < +\infty$$

Parameters. μ_Z = mean = 0

σ_Z^2 = variance = 1

Cumulative distribution function. Cumulative probabilities are found for any normal variable by first converting it to a standard normal variable as described above and using standard normal probability tables found in any basic statistics text.

Uses.

1. The distribution of measurement errors is usually normal.

2. Sample statistics such as the mean are distributed normally, which is ideal for making statistical inferences.

3. Corrective maintenance times are often normally distributed.

4. The normal is a good approximation to binomial and Poisson variables when sample sizes are large.

5. The use of the normal distribution often yields satisfactory results even when variables are not normally distributed.

EXAMPLE 7-5. A manufacturer estimates that its truck tires have a lifetime which is normally distributed with a mean of 13,000 miles and standard deviation of 2000 miles. If it guarantees to an Air Force buyer that the tires will last 10,000 miles, what percentage of tires can be expected to wear out during the guarantee period?

First it is necessary to use the transformation

$$Z = \frac{X-\mu}{\sigma} = \frac{10,000 - 13,000}{2,000} = -1.5$$

therefore

$$P(X < 10,000) = P(Z < -1.5) = 0.0668$$

As a result, the Air Force buyer can expect that 6.68% of the tires will wear out prior to the expiration of the warranty period.

Lognormal Distribution

Definition. The continuous random variable x is said to have a lognormal distribution if the log x is normally distributed and x > o.

Uses. The repair times of many equipment items, particularly electronic equipment, are often skewed right due to the extensive repair times of a few maintenance actions. In these cases, repair times often follow lognormal distributions in which the logarithms of repair times (not the repair times themselves) are normally distributed. Probabilities are calculated using the procedures for the normal distribution.

EXAMPLE 7-6. In a problem dealing with repair times of a component, the logarithm of each repair time is used to determine probabilities, instead of actual repair time. Thus, the mean and variance of the distribution can be found by:

$$\text{Mean: } \mu_{\log x} = E(\log x) = \frac{\sum_{i=1}^{N} \log x_i}{N}$$

here x_i = time to repair i^{th} item

N = number of items repaired

The mean time to repair (MTTR) can be found by taking the antilog of $\mu_{\log x}$. In essence, the mean of the lognormal distribution is the geometric mean of the actual observations of repair times.

Variance:

$$\sigma_{\log x}^{2} = \frac{\sum_{i=1}^{N} (\log x_i)^2 - \left(\sum_{i=1}^{N} \log x_i\right)^2 \Big/ N}{N}$$

Where x_i and N are the same as above.

Exponential Distribution

Definition. A continuous distribution which is a special case of the gamma distribution. If the distribution of discrete events is Poisson, the time (or interval) between the occurrences of such events will be exponentially distributed under the following conditions.

1. The medium of measurement is continuous such as time, length, or volume.

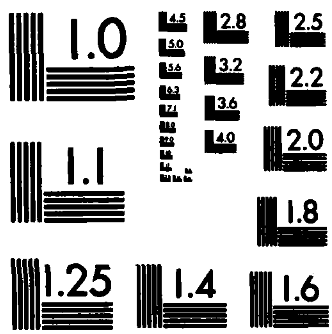
2. A subset of the medium can be defined such that the probability of two or more occurrences in the subset is zero.

3. The probability of occurrence of an event in any subset of the medium is independent of the occurrence in any other subset.

4. The probability of occurrence of the event is proportional to the size of the subset being considered.

AD-A121 990 QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS J E ENGEL APR 80 AFIT-LS-32 3/7

UNCLASSIFIED F/G 15/5 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Probability density function.

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq o$$

Cumulative distribution function.

$$F(x) = \int_o^x \lambda e^{-\lambda x} dx \quad \text{for } x \geq o$$

$$= 1 - e^{-\lambda x}$$

Parameter. λ = the average number of occurrences of an event in a specified interval of time or space.

Mean. $\mu_x = \frac{1}{\lambda}$ = average interval between occurrences of an event x.

Variance.

$$\sigma_x^2 = \frac{1}{\lambda^2} = \mu_x^2$$

Uses. The exponential is sometimes referred to as the interarrival time distribution. It is often used to represent the time to failure of electronic components and other equipment items. Here $\frac{1}{\lambda}$ would represent the mean time between failures (MTBF) of a component. In this sense, the exponential distribution is known as the constant hazard rate reliability function, i.e., the failure (hazard) rate, λ , remains constant.

EXAMPLE 7-7. From historical records it is known that the MTBF of a certain component is 80 hours. If the component is to be operated for 10 consecutive hours during the next mission, what is the probability it will not fail?

If it can be assumed that time to failure of this component is exponentially distributed, let x = time between failures.

Thus μ_x = 80 hours

and $\lambda = \frac{1}{\mu_x} = \frac{1}{80}$ hours

The probability that the component will operate for more than 10 hours is found by:

$$P\ (x > 10) = 1 - P(x \leq 10) = 1\text{-}F(10)$$

$$\text{or } 1 - \left[1\text{-}e^{-\left(\frac{1}{80}\right)(10)}\right]$$

$$= e^{-\left(\frac{1}{80}\right)(10)} = e^{-\frac{1}{8}} = 0.882$$

The chances are very good the equipment will not fail during the next mission. This probability (0.882) is really the reliability of the equipment for a 10-hour period.

CHAPTER 8

ESTIMATION AND HYPOTHESIS TESTING

"We use Reason for improving the Sciences; whereas we ought to use the Sciences for improving our Reason."

Antoine Arnaulde, 1662
(The Post-Royal Logic)

This observation by Arnaulde is certainly as applicable today as it was in the fifteenth century. It is the purpose of this chapter to explain and demonstrate the use of science (statistics, in particular, probability and hypothesis testing) for improving our reason (ability to make a decision).

INTRODUCTION

Statistical analysis has come to play a central role in the decision making process within all of the logistical disciplines. The type of analysis introduced here involves the decision to take one action or another based upon the acceptance or rejection of a statement called an hypothesis. The hypothesis can be a statement about the output of a machine, the number of defective parts in a supply shipment, the fuel flow in a turbine engine, or any other measurable event of interest to the decision maker. The hypothesis is actually a statement made about the value of a parameter. To test the hypothesis, i.e., determine whether or not to accept the value of the parameter as being true, a sample is taken. When the sample statistic differs from the parameter stated in the hypothesis, a decision must be made as to whether the difference is a consequence of random sampling error or because of a real difference between the population from which the sample was drawn and the population whose parameter is stated in the hypothesis. The latter difference is called a "significant difference" and would cause the decision maker to reject the hypothesis, i.e., to conclude that the statement about the parameter is not true. For this reason, tests of hypotheses are often referred to as tests of significance. However, prior to developing the methodology for conducting tests of hypotheses, relevant and related material on statistical estimation is presented to provide the foundation necessary for conducting such tests.

STATISTICAL ESTIMATION

Fundamental to statistical estimation is the concept of a random sample. Decisions must be made without complete information. Every grenade in storage can't be exploded in order to determine the rate of defectives. Every drop of water in a reservoir cannot be tested for suspended particulates. Obviously, this would be too costly and wasteful. However, the desired information about the population of interest, i.e., the supply

of grenades or the water in the reservoir, can be obtained by inferences derived through the examination of only a small portion of the population. In other words, information about the characteristics of any population can be obtained through a random sample.

Sampling Theory

Let X be a random variable which represents a population having a distribution with a mean μ and variance σ^2. Also let X_i be a random variable, the i^{th} value observed in a random sample from the above population. If the sample is random, then $X_1, X_2, \ldots, X_n$ are n independent random variables each having the same distribution as the population. By definition, a simple random sample is one in which every element in the population has an equal and independent chance of being selected. Further, a statistic is a function of one or more random variables from a random sample that does not depend upon any population parameters. The random sample, through the statistic which is calculated from it, becomes the vehicle for making inferences about the population of interest. The sample statistic then, is an estimator of a population parameter. Common notation is introduced below:

Name	Sample Statistic	is an estimator of	Population Parameter
Mean	$\bar{X}$ or $\hat{\mu}$		μ
Variance	s^2 or $\hat{\sigma}^2$		σ^2
Standard deviation	s or $\hat{\sigma}$		σ
Proportion	p or $\hat{\pi}$		π
Standard error of the mean	$s_{\bar{x}}$ or $\hat{\sigma}_{\bar{x}}$		$\sigma_{\bar{x}}$

Calculations

Algebraic expressions used to compute the statistics are given below along with the formulas which can be used to calculate the values of population parameters. The formulas are common to all statistics texts and are stated without proof. It should be noted that whenever a sample consists of all possible random variables in the population, then the sample is called a census and the value of the sample statistics which are computed are the actual values of the population parameters.

	Sample	Population
Mean	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$	$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$
Variance	$s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$	$\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$
Standard deviation	$s = \sqrt{s^2}$	$\sigma = \sqrt{\sigma^2}$
Standard error of the mean	$s_{\bar{x}} = \sqrt{\dfrac{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})}{n-1}}{n}}$	$\sigma_{\bar{x}} = \sqrt{\dfrac{\dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}{n}}$

Where n = number observations in sample
N = total number of observations (or values) in population

Properties of Estimators

In a previous section it was stated without proof that certain sample statistics are estimators of population parameters. The reason why $\bar{x}$, s^2, s, and p are used as estimators is because they possess certain desirable properties. Again, without proof, the desirable properties of a good estimator are given below. Consider $\hat{\theta}$ as an estimator of θ, then

1. θ is <u>unbiased</u> if $E(\hat{\theta}) = \theta$
2. θ is <u>consistent</u> if $E(\hat{\theta}) \rightarrow \theta$ and $Var(\hat{\theta}) \rightarrow 0$ as $n \rightarrow \infty$
3. If $\hat{\hat{\theta}}$ is another estimator for θ, then $\hat{\theta}$ is more <u>efficient</u> than $\hat{\hat{\theta}}$ if $Var(\hat{\theta}) < Var(\hat{\hat{\theta}})$
4. If $\hat{\theta}$ contains all relevant information relative to the parameter it is estimating, then $\hat{\theta}$ is a <u>sufficient</u> estimator of θ.

Stated without proof, $\bar{X}$, s^2, s, and p are unbiased, consistent, efficient, and sufficient estimators of μ, σ^2, σ, and π, respectively. Of course, there are other estimators of these population parameters, but they are used only under certain conditions because they do not contain the desired properties.

For example, the sample median can be used to estimate μ, the population mean. However, the sample mean $\bar{X}$ is a more efficient estimator because it has a smaller variance than the sample median. In short, the <u>best estimator</u> can be defined as the unbiased estimator having the smallest variance.

As another example, we know that s^2 is an estimator of σ^2. Recall that

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}.$$ Why divide by n-1?

Another estimator of σ^2 is S^2, where

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$$

It can be shown, however, that S is biased, i.e., $E\left[S^2\right] = \left[\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}\right] \neq \sigma^2$.

So, instead, we use s^2 as defined above.

In addition to the four properties listed above, $\bar{X}$ the sample mean has other desirable attributes.

<u>Central Limit Theorem</u>

One of the most fundamental and important theorems in inferential statistics involves the distribution of the sample mean $\bar{X}$. The <u>central limit theorem</u> states that:

> If a population has a mean μ and a finite standard deviation σ, then the distribution of the means of all possible samples of size n drawn from that population will be approximately normal with a mean μ and standard deviation $= \frac{\sigma}{\sqrt{n}}$.

It is important to note that the distribution of sample means $\bar{X}$ will approach a normal distribution even when the population itself is <u>not</u> normally distributed. The distinction should be made between the distribution

of the sample and the distribution of the sample mean. If the sample is random, then the distribution of the sample should resemble the distribution of the population. In contrast, the distribution of the sample <u>mean</u> will be normal. The understanding of this important distinction is essential to the application of sampling theory to estimation and hypothesis testing.

<u>Sampling Distributions</u>

In many cases it is convenient (and correct) to assume that a population is normally distributed. When this is true, making inferences about population parameters is a simple, straightforward task. To aid in the process of statistical estimation several distributions have been developed. Of concern here are the t, χ^2, and F because sample statistics have been shown to follow these distributions. The use of these distributions requires that:

1. the populations are approximately <u>normally</u> <u>distributed</u>, and that

2. the observed values of the random variable in the sample are <u>independent</u>. This will be true if the sample is random.

These observations are often referred to as normally independently distributed (NID) random variables. As such, they possess certain desirable characteristics. For example, the sum of NID random variables is itself normally distributed and the sum of the squares of NID random variables has a Chi-square (χ^2) distribution.

The t distribution is useful in making inferences when using $\bar{X}$ as an estimator of μ. It can be shown that

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

Like the normal distribution, the t distribution is continuous, symmetric, and unbounded. Its shape depends upon its degrees of freedom. As degrees of freedom get larger, the t approaches a normal distribution. Degrees of freedom for the t distribution are equal to the sample size minus one (n-1).

The χ^2 distribution is useful in making inferences about σ^2 when s^2 is used as an estimator. The χ^2 distribution has degrees of freedom (ν) equal to n-1 and is related to the t distribution in the following manner. Recall

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

When the denominator is multiplied by $\frac{\sigma}{\sigma} = 1$ then

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}} \cdot \frac{\sigma}{\sigma}} = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\frac{s}{\sigma}} .$$

The numerator, $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ is the standard normal variable $Z \sim N(0,1)$, and the

denominator $\frac{s}{\sigma} \sim \sqrt{\frac{\chi^2_\nu}{n-1}}$, or the χ^2 distribution divided by its degrees of freedom. Squaring and rearranging this expression results in

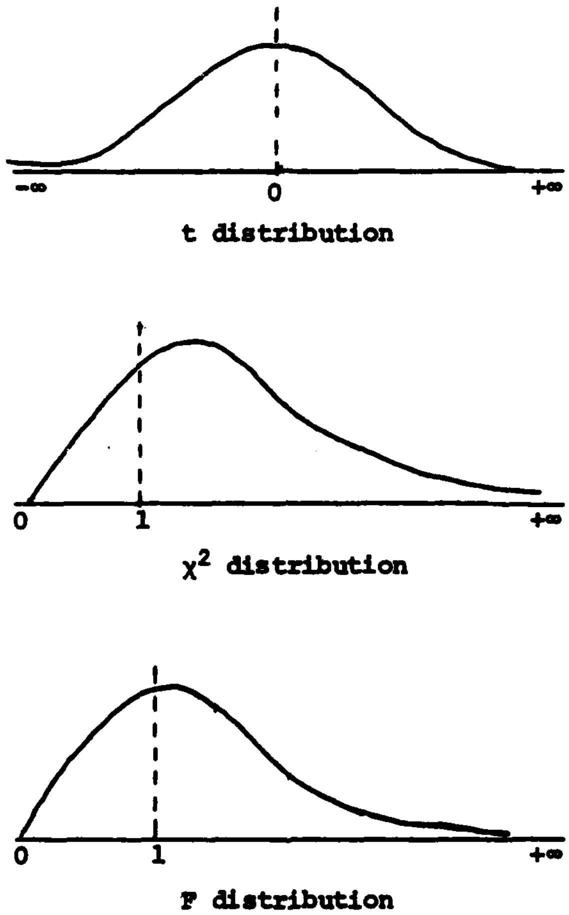
$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_\nu$$

where $\nu = n-1$. The χ^2 is a continuous distribution which has a lower bound of zero and no upper bound. As such, it is skewed right. (See Figure 8-1). Its shape depends upon ν.

The F distribution is the ratio of two random variables, each having a χ^2 distribution divided by their respective degrees of freedom. Therefore,

$$F^{\nu_1}_{\nu_2} = \frac{\frac{\chi^2_{\nu_1}}{\nu^1}}{\frac{\chi^2_{\nu_2}}{\nu_2}}$$

Figure 8-1



Like the χ^2 distribution, the F distribution is continuous has a lower bound of zero and no upper bound, and is skewed right. It is related to the t distribution as

$$t_\nu^2 = \frac{[N(0,1)]^2}{\frac{\chi_\nu^2}{\nu}} = \frac{\chi_\nu^2 = 1}{\chi_\nu^2} = F_\nu^1$$

thus, t_ν^2 is the same as the F distribution with one degree of freedom in the numerator and ν degrees of freedom in the denominator.

In summary, the use of sampling distributions make it very convenient to make inferences about the population parameters μ and σ^2. In later sections these distributions will be used to test hypotheses and construct confidence intervals for these parameters. The t distribution will be used to infer on μ; the χ^2 and F distributions will be used to infer on σ^2.

Point Estimation

Two kinds of estimation are useful. The first is called a point estimate wherein a single value is used to estimate a population parameter. Point estimators have already been discussed. Thus, if it were desired to estimate the average useful life of Air Force bus tires, a sample of tires would be selected from the population of all Air Force bus tires. The sample mean $\bar{X}$ of miles before wearout would be computed. This would then be the best estimate of the average miles before wearout of all Air Force bus tires. The estimate could be used for planning new tire purchases.

Note that $\bar{X}$ is the mean of a single sample of tires in the illustration. If another sample were taken, the new $\bar{X}$ would most likely differ from the first. The question to be asked would be, which one is the correct $\bar{X}$ to use? By definition, both are "best estimators" of μ. The answer lies in the use of an interval estimator.

Interval Estimation

The objective of interval estimation is to provide a range of values in which the population parameter is estimated to lie. Two values are specified, the upper and lower bounds. Along with the bounds, a degree or level of confidence is specified which can be defined as the probability with which the population parameter will be "captured" by the two bounds. The resulting interval is called a confidence interval.

The level of confidence is expressed as

$$100\ (1 - \alpha)\ \%$$

where α is some small value less than 1, such as 0.05 or 0.01, which represents the probability that the interval will not capture the population parameter.

In developing estimates for the population mean μ, two distributions are used, the normal and the student t. Both of these are continuous and unbounded, i.e., the range is $-\infty$ to $+\infty$. For this reason, we can never be 100% confident that our interval will capture the population parameter. However, in the effort to place as much "confidence" as possible in our estimate, we use small values of α.

The confidence interval on the population mean is given by the following:

$$\bar{X} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Where

$\bar{X}$ = sample mean

σ = known population standard deviation

n = sample size

$Z_{\frac{\alpha}{2}}$ = value of the standard normal deviate for a specified level of $\frac{\alpha}{2}$.

This formulation assumes:

1. σ is known.
2. The sample is large, i.e., $n \geq 30$ <u>or</u> the sample was drawn from a normal population.

If, however, σ is not known, the value of Z is replaced with the appropriate value of the student t with n-1 degrees of freedom. The value of s is used to estimate the unknown value of σ. The $(1 - \alpha)\cdot 100\%$ confidence interval then becomes:

$$\bar{X} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

where $t_{\frac{\alpha}{2}}$ is the value from the student t distribution with n-1 df such that $P(t > t_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$. As in the confidence interval for μ where σ is known $\frac{\alpha}{2}$ is used to indicate a "two-sided" interval. In the case where a

"one-sided" confidence interval is desired, use:

$$\bar{X} - t_{\alpha} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \infty \quad \text{upper or right-side interval}$$

or

$$\bar{X} + t_{\alpha} \cdot \frac{s}{\sqrt{n}} \geq \mu \geq -\infty \quad \text{lower or left-side interval}$$

The probability statement for a lower confidence interval for μ can then be constructed as:

$$P(\mu \leq \bar{X} + t_{\alpha} \cdot \frac{s}{\sqrt{n}}) = 1 - \alpha$$

To illustrate, consider the case where contract specifications require that the tensile strength of a certain item on the average be no less than 200 psi. Since destructive testing is costly, the quality control engineer decides to test a small sample of 15 items and obtains the following results.

$\bar{X}$ = 254 psi	n = 15 items
s = 24 psi	$t_{\alpha = .01}$ = 2.624 (degrees of freedom = 14)

He constructs a one-sided confidence interval (upper) with a confidence level $(1 - \alpha) = .99$ in the following manner, assuming σ is unknown.

$$99\%\ CI = 254 - 2.624 \cdot \frac{24}{\sqrt{15}} \leq \mu \leq \infty = 237.7 \leq \mu \leq \infty$$

The engineer can be 99% confident that the population of items produced has an average tensile strength greater than 237.7 psi. In other words, he can be fairly certain the items are exceeding the tensile strength specifications called for in the contract.

Finite Population Sampling. When sampling from a finite population, it can be shown that Var $(\bar{X}) \neq \frac{\sigma^2}{n}$, therefore, the use of a correction factor is required. The variance of the same mean then becomes:

$$\text{Var }(\bar{X}) = \frac{\sigma^2}{n} \cdot \frac{N - n}{N - 1}$$

where N = size of the population

n = size of the sample

It can be readily seen that when the entire population is included in the sample, the variance of the sample mean is zero because:

$$\text{Var }(\bar{X}) = \frac{\sigma^2}{n} \cdot \left[\left(\frac{N - N}{N - 1} \right) = 0 \right] = 0$$

This is consistent with sampling theory in that when the entire population is included in the sample, the sample mean $\bar{X}$ is really the population mean μ, which has no variance, i.e., it is a constant. It should also be carefully noted that the variance of the sample mean gets smaller and smaller as the sample size n approaches the population size of a finite population, or when the sample size gets extremely large from an infinite population. So,

as $n \rightarrow N$, then $\frac{\sigma^2}{n} \rightarrow 0$ (finite population)

and $n \rightarrow \infty$, then $\frac{\sigma^2}{n} \rightarrow 0$ (infinite population)

Sample Size Determination. It is often desirable to develop a confidence interval at some specified width. For example, the civil engineer may wish to insure that the average weight for bags of cement delivered to his facility does not vary from μ by more than five pounds. How large a sample must be weighed? His specified conditions state that the maximum deviation E, or the allowable error of $\bar{X}$ from μ is:

$$E = \bar{X} - \mu = 5$$

The engineer wishes to be 95% confident that $\bar{X}$ does not differ from μ my more than 5 pounds. For this level of confidence then

$$E = 5 = 1.96\sigma_{\bar{X}}$$

Assuming the weights of the bags of cement are normally distributed and σ is known to be 10 pounds:

$$5 = 1.96 \cdot \frac{10}{\sqrt{n}}$$

The sample size is found by solving the equation for n.

$$(5)^2 = \left[1.96 \cdot \frac{10}{\sqrt{n}}\right]^2$$

$$n = \frac{3.84 \cdot 1000}{25}$$

$$n = 154$$

Thus, to insure 95% confidence that the sample mean will not differ from the true population mean by more than five pounds, the civil engineer will have to weigh 154 bags of cement.

The general expression for determining sample size is given by

$$n = \left[\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{E}\right]^2$$

where

n = required sample size

$Z_{\frac{\alpha}{2}}$ = value of the standard normal variable for a confidence level of $(1 - \alpha) \cdot 100\%$

σ = known standard deviation of the population. If σ is unknown, a pilot sample should be taken and s computed as an estimator

E = maximum allowable deviation from μ, or the confidence interval half-width.

In a similar manner, the appropriate sample size can be computed when a desired level of precision is necessary in estimating a population proportion.

$$n = \pi(1-\pi) \left[\frac{Z_{\frac{\alpha}{2}}}{E} \right]^2$$

It can be seen that the maximum sample size will occur when a value of 0.50 is used for π.

<u>Confidence Intervals for Variances</u>. In a manner similar to that used for developing interval estimates on μ, confidence intervals can be constructed to estimate population variances, σ^2. Since the sampling distribution for variance is χ^2_ν, and $\chi^2_\nu = \frac{(n-1)s^2}{\sigma^2}$, then a two-sided confidence interval for σ^2 can be expressed as:

$$\frac{(n-1)s^2}{\chi^2_{\nu,\,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\nu,\,\frac{\alpha}{2}}}$$

where s^2 = sample variance

$\chi^2_{\nu,\,1-\frac{\alpha}{2}}$ = upper value of the χ^2 distribution with ν degrees of freedom

$\chi^2_{\nu,\,\frac{\alpha}{2}}$ = lower value of the χ^2 distribution with ν degrees of freedom

For example, a loadmaster wished to be 95% confident of the variance in the weight of C-141 cargo and obtained the following information from a sample.

$$n = 25 \qquad s^2 = 18 \text{ (hundreds of pounds)}^2$$

The 95% confidence interval would be:

$$\frac{(25 - 1)(18)}{39.4} \leq \sigma \leq \frac{(25 - 1)(18)}{12.4}$$

or $$10.96 \leq \sigma^2 \leq 34.84$$

Similarly, to develop one-sided confidence intervals, the following can be used:

For the lower (left) side:

$$0 \leq \sigma^2 \leq \frac{(n - 1)s^2}{\chi^2_{\nu,\alpha}}$$

(Note that $\frac{\alpha}{2}$ is not used)

For the upper (right) side:

$$\frac{(n - 1)s^2}{\chi^2_{\nu,\, 1 - \alpha}} \leq \sigma^2 \leq \infty$$

In the previous example, the two one-sided intervals would be:

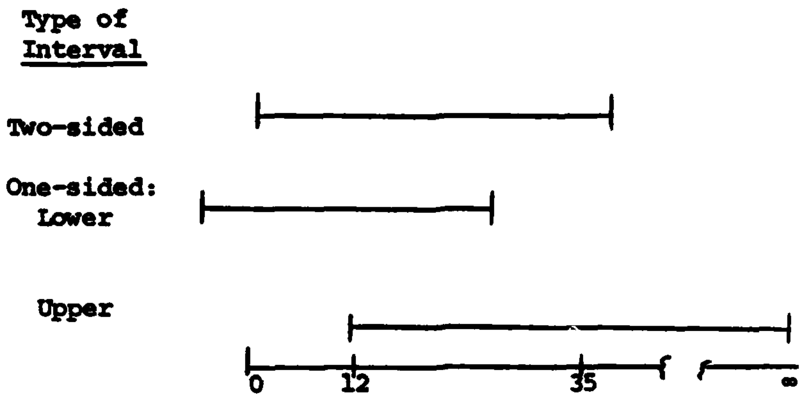
Lower: $\sigma^2 \leq \frac{(25 - 1)(18)}{13.8}$ or $\sigma^2 \leq 31.3$

Upper: $\sigma^2 \geq \frac{(25 - 1)(18)}{36.4}$ or $\sigma^2 \geq 11.87$

The appropriate values of the χ^2 distribution are obtained from the χ^2 table by finding the degrees of freedom and specifying the level of confidence desired.

Figure 8-2 graphically illustrates the relationships among the intervals computed above.

Figure 8-2



It will be shown later that confidence intervals are useful in the testing of hypotheses, a subject to which we now turn.

Statement of Hypothesis

As indicated at the beginning of the chapter, a hypothesis is a statement about a population parameter. In hypothesis testing, two hypotheses are constructed. The first is called the null hypothesis which represents a claim or a statement which we wish to refute. The other is called the alternate hypothesis which is a statement which we desire to support with the evidence supplied through a sample. It is important to note that hypotheses are always constructed on population parameters, not sample statistics.

To illustrate, consider the case of a supply officer who must accept a shipment of repair parts from a supplier unless it can be shown that the parts do not meet specifications. Suppose the specifications require that the average weight of the parts does not exceed 25 pounds. The supply officer would construct the following hypotheses.

Null hypothesis -- H_0: $\mu \leq 25$ lbs

Alternate hypothesis -- H_1: $\mu > 25$ lbs

Under the circumstances, the supply officer will accept the shipment of repair parts unless he/she can show that a sample of repair parts weighs significantly more than 25 pounds. In other words, he/she must be able to reject the null hypothesis before he/she can reject the shipment. By rejecting the null hypothesis, he/she is stating that he/she has sample evidence to support the alternate hypothesis. If he/she cannot reject the null hypothesis, it is because the sample data are not sufficient to show that the specifications are not met.

So, in any test of hypotheses, a "strong" conclusion is reached only when the null hypothesis can be rejected. If Ho cannot be rejected, the only conclusion which can be reached is that the sample data do not support the alternate hypothesis. This is known as the "weak" conclusion; weak because it is a fall-back conclusion, i.e., the sample evidence cannot support the contrary.

Level of Risk

It is possible that our supply officer will reject the null hypothesis and, consequently, the shipment of repair parts erroneously. This could be costly both in a monetary sense (contractor sues because the parts do meet specifications) and in mission sense (parts are needed to fill Not Operationally Ready--Supply requisitions). To preclude this from happening too frequently, a level of significance is established. The level of significance, called α, is set at a level which corresponds to the consequences which will be incurred if the null hypothesis is rejected when it is in fact true. When the effect of a wrong decision may involve human lives, α might be set as low as .0001. On the other hand, a hypothesis test involving the behavioral patterns of experimental rabbits, an acceptable level of α could be .25. In any event, the logistician must be cognizant of the consequences (costs) involved in making a wrong decision.

In probabilistic terms, the level of significance α, can be stated as the probability that Ho will be rejected when it is true. α then is referred to as the level of significance of the test.

Conducting Tests of Hypotheses

Four tests of hypotheses are discussed including the t tests on means and proportions, and the chi-square and F tests on variance. Each test performed can be classified as a one- or two-tailed test. For a one-tailed test, the experimenter has a priori knowledge about the direction of the difference to be ascertained through the sample. This is demonstrated in the following example.

A supplier of steel rods to the Air Force claims that their average tensile strength is 750 pounds. To test this claim, the Air Force using activity would establish the following hypotheses:

$$H_0: \mu \geq 750 \text{ pounds}$$

$$H_1: \mu < 750 \text{ pounds}$$

If the Air Force can show through a sample of rods that the tensile strength is significantly less than 750 pounds, i.e., reject H_o, then the supplier's claim has been refuted and shipments of the rods can be refused.

In the same example, the supplier would establish the following hypotheses:

$$H_o: \mu \leq 750 \text{ pounds}$$

$$H_1: \mu > 750 \text{ pounds}$$

In this case, the supplier would like to show that the rods have a tensile strength of at least 750 pounds. Obviously with same sample, only one rejection of H_o can occur. The first test is known as a one-tailed test to the left, identified by the direction of the inequality of the alternate hypothesis. The strong conclusion is reached by rejecting the null hypothesis and accepting the alternate. This conclusion is supported by sample data.

In the second case, a one-tailed test to the right, is one in which the supplier would like to reject H_o, i.e., support his claim with sample evidence.

A third situation may exist in which the direction of the test is unimportant or the experimenter has no a priori knowledge of the direction of the difference being tested.

For example, consider the case of the diameter of ball bearings. To work properly, they can neither be too large nor too small. To test, the following two-tailed hypotheses would be established:

$$H_o: \mu = 15 \text{ mm}$$

$$H_1: \mu \neq 15 \text{ mm}$$

Here, specifications would state that the diameter of ball bearings must be 15 millimeters. Rejection of H_o, i.e., acceptance of the strong conclusion would occur if a sample of ball bearings had average diameters either significantly less than or significantly greater than 15 mm.

<u>t Tests on Means (one sample)</u>

Two-tailed Test: $H_o: \mu = \mu_o$
$H_1: \mu \neq \mu_o$

One-tailed Test: $H_o: \mu \leq \mu_o$
$H_1: \mu > \mu_o$

or

$H_o: \mu \geq \mu_o$
$H_1: \mu < \mu_o$

μ_o is a value to be tested

Assumptions:

1. Random sample
2. Normal population or sample size $\geq$ 25
3. σ^2 unknown

Test statistic:*

$$t = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}}$$

where
$\bar{x}$ = sample mean

μ_o = hypothesized value to be tested

s = sample standard deviation

n = sample size

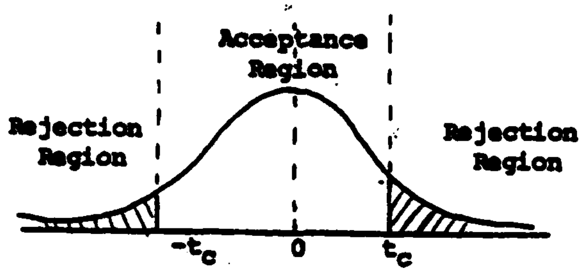
Decision rule: Reject H_o if t is more extreme than t_c, a critical value of t determined by the level of α chosen and degrees of freedom of the sample $(n - 1)$.

*When σ^2 is known, the standard normal variable is used as the test statistic, where

$$Z = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}}$$

Z is compared with Z_c, from the standard normal table at the specified level of α.

Illustrated for a two-tailed test:



or $\quad -Z_c \quad\quad Z_c$

If the test statistic falls into the cross-batched area, reject H_o and accept H_1.

t Tests on Difference Between Means (two sample)

Tests of means for two samples encompass four separate cases. Hypotheses, however, are similar in each case.

Two-tailed test: H_o: $\mu_1 = \mu_2$ $\quad\quad$ One-tailed test: H_o: $\mu_1 \leq \mu_2$

H_1: $\mu_1 \neq \mu_2$ $\quad\quad$ H_o: $\mu_1 > \mu_2$

or $\quad H_o$: $\mu_1 - \mu_2 = 0$ $\quad\quad$ or $\quad H_o$: $\mu_1 \geq \mu_2$

H_1: $\mu_1 - \mu_2 \neq 0$ $\quad\quad$ H_1: $\mu_1 < \mu_2$

Case 1: Paired Data--nonindependent samples, sometimes referred to as the "before and after" test.

Assumptions:

1. Paired data (related samples)
2. σ^2 unknown
3. σ^2 remains constant for both samples.
4. Normal population or sample size ≥ 25.

Test statistic:

Let $D_i = X_{1i} - X_{2i}$

then $t = \frac{\bar{D}}{s_{\bar{D}}}$ where $s_{\bar{D}} = \frac{s_D}{\sqrt{n}}$

$$\text{and } s_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \bar{D})^2}{n-1}}$$

and t has n-1 degrees of freedom.

Example. This test could be used to measure the effectiveness of a training program where the <u>same</u> individuals are tested <u>before</u> training and <u>after</u> training. A significant improvement in test scores after training would indicate that the program was effective.

Case 2: Independent Samples

Assumptions:

1. Independent samples
2. Normal populations or sample sizes $\geq$ 25
3. σ_1^2 and σ_2^2 known

Test statistic:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

This test along with Cases 3 and 4 can be used to determine whether two independent samples were drawn from the same population, i.e., with the same mean.

Case 3: Independent Samples

Assumptions:

1. Independent samples

2. Normal population or sample sizes $\geq$ 25

3. σ_1^2 and σ_2^2 are unknown but equal (see F test on variance)

Test statistic:

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{and } s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$$

and t has $n_1 + n_2 - 2$ degrees of freedom

Case 4: Independent Samples

Assumptions:

1. Independent samples

2. Normal population or sample sizes $\geq$ 25

3. σ_1^2 and σ_2^2 are unknown and <u>not</u> equal (see F test on variances)

Test statistic:

$$t^* = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

and t* has an approximate t distribution with

$$\text{degrees of freedom} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2}}$$

t tests on Proportions

One sample:

Two-tailed test: $H_0: \pi = \pi_0$

$H_1: \pi \neq \pi_0$

One-tailed test: $H_0: \pi \leq \pi_0$

$H_1: \pi > \pi_0$

or $H_0: \pi \geq \pi_0$

$H_1: \pi < \pi_0$

Test statistic:

$$Z = \frac{p - \pi_0}{\sigma_p}$$

where

p = sample proportion or the number in the sample having a specific attribute divided by the total number of observations in the sample

π_0 = the hypothesized proportion to be tested

σ_p = standard error of the proportion which can be found by

$$\sigma_p = \sqrt{\frac{\pi_0 (1 - \pi_0)}{n - 1}}$$

This test statistic should be used only when the sample is large, i.e., when $\pi_o \cdot n \geq 25$. Otherwise the binomial distribution should be used to test the hypothesis.

Example. A clothing sales outlet has found that 15% of new short-sleeved shirts are returned by AF customers because they are missized by the manufacturer. The maker was advised of the problem and subsequently developed a new sizing process. Of the next batch of 500 shirts sold, 60 were returned because of missizing. Was the new process effective in reducing the number of returns?

$$H_o: \pi \geq .15$$

$$H_1: \pi < .15$$

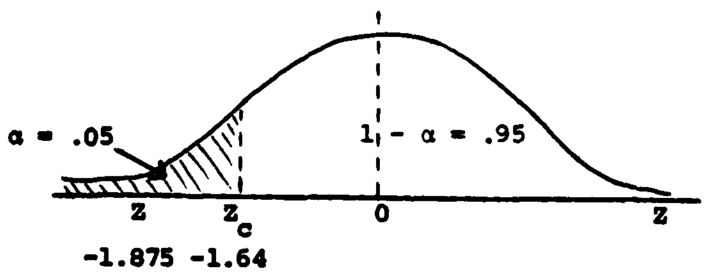
test at $\alpha = .05$

$$p = \frac{60}{500} = .12$$

$$\sigma_p = \sqrt{\frac{(.15)(1 - .15)}{500 - 1}} = 0.016$$

$$Z = \frac{.12 - .15}{.016} = -1.875$$

$$Z_c = -1.64$$



Reject H_o since $Z < Z_c$ and accept H_1 , the new process significantly reduced the proportion of defectives (returns for missizing).

Tests on Variances (One Sample)

Two-tailed test: $H_o: \sigma^2 = \sigma_o^2$ One-tailed test: $H_o: \sigma^2 \leq \sigma_o^2$

$H_1: \sigma^2 \neq \sigma_o^2$ $H_1: \sigma^2 > \sigma_o^2$

or $H_o: \sigma^2 \geq \sigma_o^2$

$H_1: \sigma^2 < \sigma_o^2$

Assumption: All tests of variances assume that samples have been drawn from normal or near normal populations.

Test statistic: One sample tests of variance involve use of the Chi-square statistic,

where,

$$\chi_s^2 = \frac{s^2(n-1)}{\sigma_o^2}$$

where s^2 = sample variance

n = sample size

σ_o^2 = hypothesized value of the variance to be tested

The χ_s^2 value, test statistic, has n-1 degrees of freedom.

Decision rule: Reject H_o if $\chi_s^2 > \chi_c^2$, where χ_c^2 is the critical value of the χ^2 with n-1 degrees of freedom at a specified level of α, obtained from a table of χ^2 values.

Example. The base supply officer will discontinue doing business with a supplier whose delivery schedule varies by more than 10 $(\text{days})^2$, i.e., $\sigma^2 > 10$. He records the delivery times of Supplier A for the next 25 deliveries and finds $s^2 = 18$, i.e.,

$$\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} = 18.$$

where X_i = number of days to deliver the i^{th} order

n = number of orders

$\bar{X}$ = average number of days to deliver an order

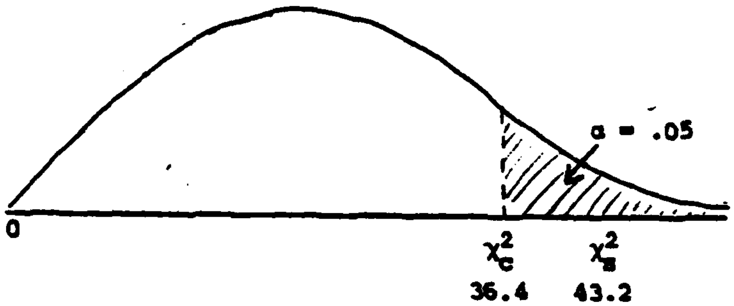
The test is performed at $\alpha = .05$

$$H_o: \sigma^2 \leq 10$$

$$H_1: \sigma^2 > 10$$

$$\chi_s^2 = \frac{(18)(25 - 1)}{10} = 43.2$$

$$\chi_c^2 = 36.4$$



Reject H_o, $\chi_s^2 > \chi_c^2$ and conclude the variability of delivery times from Supplier A exceeds 10 $(\text{days})^2$.

An alternate method for computing a test statistic can be used to find the largest acceptable value of s^2, which can be called s_c^2, where

$$s_c^2 = \frac{\chi_c^2 \cdot \sigma_o^2}{n - 1}$$

where χ_c^2 is the critical value of χ^2 with n-1 degrees of freedom at a specified level of α. In this case, reject H_o when $s^2 > s_c^2$.

In the example,

$$s_c^2 = \frac{(36.4)(10)}{(25 - 1)} = 15.2$$

Since $s^2 > s_c^2$, reject H_o, the conclusion is the same as above.

Tests of Variance (Two Sample)

Recalling the t tests on differences of means of two independent samples, it is necessary in deciding the appropriate case to apply, to determine whether unknown variances are equal (Case 3) or unequal (Case 4). The appropriate procedure is to perform an hypothesis test on the sample variances in question. To do this, an F test is applied. Again, it must be assumed that the independent samples are drawn from normal or near normal populations.

Two-tailed tests: $H_o: \sigma_1^2 = \sigma_2^2$ One tailed test: $H_o: \sigma_1^2 \leq \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$

or $H_o: \sigma_1^2 \geq \sigma_2^2$

$H_1: \sigma_1^2 < \sigma_2^2$

Test Statistic:

$$F = \frac{s_1^2}{s_2^2}$$

where s_1^2 = variance computed from the sample with the <u>largest</u> variance

s_2^2 = variance from the sample with the <u>smallest</u> variance

F has a distribution with $n_1 - 1$ degrees of freedom in the numerator and $n_2 - 1$ degrees of freedom in the denominator.

Decision rule: Reject H_o if $F^{(n_1-1)}_{(n_2-1)} > F_c{}^{(n_1-1)}_{(n_2-2)}$

<u>Example</u>. The weather service is testing the variance of bursting strength of balloons from two competitors and has gathered the following data.

<u>Company A</u>	<u>Company B</u>
$n_A = 21$	$n_B = 16$
$s_A = 2.5$ psi	$s_B = 1.5$ psi

Is there a significant difference in the variance of bursting strength between the two companies' balloons? The test is to be performed at $\alpha = .02$.

$$H_0: \sigma_A^2 = \sigma_B^2$$

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

$$F = \frac{(2.5)^2}{(1.5)^2} = 2.778$$

$$F_{c(16-1),\ \frac{\alpha}{2} = .01}^{(21-1)} = 3.37$$

$F \not> F_c$, do not reject H_0. Therefore we cannot conclude there is a significant difference in the variances of bursting strength between the two companies' balloons. Note that in a two-tailed test, the value of α is halved.

In summary, to conduct tests of hypotheses, the following methodology is appropriate:

1. State the hypotheses. It is often best to establish the alternate hypothesis first. H_1 normally contains a statement of that which is to "proved," or supported by sample data. The null hypothesis then includes all other possibilities. In many cases, the null hypothesis is a manufacturer's claim. Here it is assumed that the claim will be accepted unless it can be shown (by sample evidence) that it is not true. Decide whether the test is to be one- or two-tailed .

2. Set the level of significance. What are the costs of being wrong, i.e., rejecting a true hypothesis? For two-tailed tests, use $\frac{\alpha}{2}$.

3. Select the appropriate test statistic. The choice of the test statistic will depend upon the statement made in the null hypothesis.

4. Select the sample and compute the test statistic.

5. Compare the computed value with the critical value obtained from the appropriate statistical table.

6. Reject H_0 if the test statistic is more extreme than the critical value. Otherwise, do not reject H_0.

7. Draw the conclusion. State what has been accomplished by rejecting (or not rejecting) the null hypothesis.

CHAPTER 9

INDEX NUMBERS

Index numbers are ratios, usually expressed as percentages, indicating changes in values, quantities or prices. Typically, the changes are measured over time, each index being compared with the corresponding index from some selected base period. Simple index numbers deal with homogeneous denominations representing commodities such as plywood, steel or grain. More commonly, though, index numbers are aggregates of a number of different commodities, products or services. Each item in the aggregation is weighted in proportion to its amount in a particular end item, industry or geographical region.

Index numbers are commonly classified into three different types: price, quantity, and value. A series of price index numbers represent changes in prices of items or commodities over time. An example of a price index number series is the Wholesale Price Index (published monthly by the Bureau of Labor Statistics) which represents the change, over time, in the average wholesale prices of commodities and products sold in the United States. A quantity index measures the change in the amount of a commodity or product produced over periods of time. The Federal Reserve Board compiles a quantity index called the Index of Industrial Production which measures physical volume of factory production in the United States from one year to the next. A value index combines changes in both price and quantity over time. Value indexes can be considered to be the product of a price index and quantity index. A commonly used value index is the Index of Retail Sales published in the Federal Reserve Bulletin which reflects the changes in both the prices and the quantities of items sold by retail sales outlets across the United States.

When dealing with index numbers, it is important to identify the type of index number one is working with and use that type of index number consistently throughout the analysis. In the following discussion of the different techniques used to construct index numbers, the equations and examples shown are for price indexes. However, a simple renaming of the variables would make the equations appropriate for any type of index number construction.

Constructing Price Index Numbers

Price index numbers indicate price changes with respect to time for some specific commodity, product or service. As historical indicators, index numbers become more accurate if they are constructed using actual prices paid for a particular commodity, product or service rather than than using the more general aggregative index numbers published by agencies such as the Bureau of Labor Statistics. Accordingly, this section will treat five methods of price index number construction including examples. They are: simple price indexes, simple aggregative price indexes, weighted aggregative price indexes of the Laspeyres and Paasche types, and weighted average of relative indexes. A short discussion of selecting the base period is included before presenting the individual methods of construction.

The Base Period

Price index numbers are price relatives, usually expressed as percentages. As price relatives, they relate prices paid in one time period to prices paid in some base time period. To provide comparability, a series of index numbers representing some commodity, product or service is always constructed using the same base period, thus reflecting a percentage increase or decrease in prices relative to that base period.

The selection of a base period is usually an arbitrary process. On a short series of data, say five to ten years, the analyst often chooses the first (earliest) year as the base year. Under ideal conditions, it is best to choose as a base year a year in which prices are not changing erratically. This is difficult when hundreds of items are included in an aggregative index number and leads one back to an arbitrary choice of a base year.

A shift in a series of index numbers to a new base year is a straightforward calculation if it is known that the original sample of goods and services remains representative of the price changes to be indicated. For example, consider the problem of adjusting a 1976 index number with a base year of 1967 to a base year of 1972. In addition to the value of the 1976 number (say 151.9) one needs the value of the corresponding number in the same series in 1972 (say 118.0). The conversion is accomplished by dividing the index number to be changed by its corresponding value from the new base year and expressing the result as percentage. In this example, the 1976 index number with a new base year of 1972 would be ($100 \times 151.9 \div 118.0$) or 128.7.

The same information is relayed by the revised index number series as the initial series. Only the base year has been changed.

Simple Price Indexes

A simple price index series is usually a time series of price relatives converted to percentages. A price relative in this case is the average price of an item for a given time period (e.g., a given year) divided by the average price of an item for the base time period.

In constructing price index numbers, it is important to express a price in dollars per measure of quantity (e.g., \$/#, \$/person or $\$/ft.^2$). These measures of quantity are used in the construction of the weighting factors for weighted aggregative index numbers. One should not express the price in terms of dollars per period such as it found in accounting data. Accounting data need to be edited to dollars per measure of quantity before their use in price index number construction.

There are four steps to be followed in constructing a simple price index number. They are: (1) Collect or develop average time series price data for the product, commodity or service to be analyzed. For example, assume the average yearly prices of ¼ inch thick, 4 feet by 8 feet, grade AC, interior, sanded plywood per 1000 square feet are:

Year	1976	1969	1971	1973	1975
Price	\$84.12	\$95.06	\$107.32	\$121.35	\$152.72

(2) Select a base year, say 1967; (3) Calculate a time series price relative for each year by dividing each price by the base year price; and (4) Convert the series to an index number of percentage.

Year	1967	1969	1971	1973	1975
Price Relative	1.000	1.130	1.276	1.443	1.816
Index Number	100.0	113.0	127.6	144.3	181.6

These index numbers indicate the percentage change in price with respect to the base year only. The index for 1971 of 127.6 indicates that the average price of plywood went up 27.6 percent with respect to 1967. It does not indicate that the average increase was 14.6 percent (127.6 - 11.30) with respect to 1969. To calculate the percentage increase in price for 1971 with respect to 1969 one would divide the 1969 index into the 1971 index, multiply the dividend by 100 and subtract 100.0.

$$100 \left[(127.6 \div 113.0)\right] - 100.0 \text{ or } 12.9 \text{ percent}$$

Mathematical notation is useful in expressing different methods of index number construction. For constructing simple price indexes, the notation is

$$I_j = 100 \left[\frac{P_j}{P_o} \right]_o^n$$

where j is a subscript denoting different periods ranging from o to n. In this instance P_o is the price during the base year 0. P_j are the price values for the different years, e.g., P_o = \$84.12 and P_2 = \$95.06. I_j are the indexes for the different years, e.g., I_o = 100.0, and I_2 = 113.0.

Seldom will a single simple index number suffice for pricing purposes. Most items that are purchased are made up of many different materials and types of labor, the prices of which vary at different rates as time proceeds. Therefore, the analyst must construct composite index numbers that reflect aggregative changes in the prices of the components, assemblies, and types of labor that make up an item. This need has been satisfied by developing a number of different methods for constructing aggregative indexes. The easiest of these methods is the simple average of relatives approach.

Average of Relatives Price Indexes

Simple average of relatives price indexes are constructed to indicate time series changes in prices of more than one item or commodity. Consider the construction of wooden boxes from plywood and nails. One can construct simple price indexes for nails in the same fashion that the indexes for plywood were constructed in a previous example.

Year	1967	1969	1971	1973	1975
Price per keg	\$15.20	\$15.65	\$16.00	\$15.92	\$18.10
Price relative	1.000	1.030	1.053	1.047	1.191
Index number	100.0	103.0	105.3	104.7	119.1

To construct a simple average of relatives price index reflecting an aggregative price change for plywood and nails, one sums the indexes for each commodity during a year and divides by the number of commodities. In this instance, with only two indexes, the calculations are easy. For example, in 1969 the index is (113.0 + 103.0) ÷ 2 or 108.0.

Year	1967	1969	1971	1973	1975
Simple Aggregative Price Index for Plywood & Nails	100.0	108.0	116.5	124.5	150.4

Continuing the practice of writing equations for index number construction, the general relationship for the simple average of relative price index is:

$$I_j = \sum_{i=1}^{m} \frac{I_{ij}}{m} = \frac{100 \sum_{i=1}^{m} \frac{P_{ij}}{P_{io}}}{m}$$

where i = 1 to m and relates to the different commodities, products or services that make up the Index I.

and j = 0 to n and relates to the different periods from the first period 0 to period n.

An advantage of the simple average of relatives price index is that it permits the user to mix products with different dimensions. In this illustration one product with dimensions of 1000 ft^2 is mixed with another product with dimensions of kegs. This is done by reducing each product to a dimensionless ratio before combining.

A disadvantage of the simple aggretative pricing index method is that it implicitly assigns weights to each of the procucts or commodities. Thus, in the illustration, 1000 ft^2 of plywood has the same weight as one keg of nails. This weighting would be acceptable only if in constructing boxes, one keg of nails was needed for each 1000 ft^2 of plywood used. Seldom would that circumstance occur. Thus, there is a need to weight each price relative in constructing an aggregative index number.

Choosing Weighting Factors

The quantities of each item or product in an aggregative index number are the logical candidates for explicit weighting factors. There is much discussion in the literature concerning which quantities to use, base year, current year or some mixture of each.

Base year quantities are the quantities of each item in the index purchased in the base year. In base year quantity construction, these quantities would be used to weight price relatives for all years of the time series, thus eliminating the effects of quantity changes from one year to another on the price index number. This approach does not recognize that in the real world the mix of items being purchased often changes from one year to another. Base year weighting is used in constructing price relatives of the Laspeyres type. Another approach, current year weighting, is used in constructing aggregative price indexes of the Paasche type.

Weighted Aggregative Price Indexes of the Laspeyres and the Paasche Types

A common approach for index number construction is the fixed base year weighting approach developed by Etieene Laspeyres in 1864. In addition to collecting time series price data for the different items to be aggregated, the approach requires collection of quantity data for the base year selected. The formulation of the Laspeyres methods is:

$$I_j = 100 \left[\frac{\sum_{i=1}^{m} P_{ij} \; Q_{io}}{\sum_{i=1}^{m} P_{io} \; Q_{io}} \right]$$

Where i stands for the different items numbered from 1 to m and j stands for the different periods, j=o is the base period or year, thus P_{io} stands for the base year price of the i^{th} commodity and Q_{io} stands for the base year quantity of the i^{th} commodity.

Extending the plywood and nails example, one needs to collect quantity data as well as price data. Summarizing the data again:

Year	1967	1969	1971	1973	1975
Plywood $\frac{\text{Price}}{\text{M}}$	\$84.12	\$95.06	\$107.32	\$121.35	\$152.72
Plywood Qty. M	1000	1025	1025	1010	1010
Nails $\frac{\text{Price}}{\text{keg}}$	\$15.20	\$15.65	\$16.00	\$15.92	\$18.10
Nails Qty. Kegs	102	104	105	103	99

Note that the quantities are expressed in the same dimensions as the prices, e.g., plywood in 1000 ft^2 or M and nails in kegs. For the Laspeyres method of construction, only the base year (1967) quantities will be used for weighting purposes. For example:

$$I_{1975} = 100 \left[\frac{\sum_{i=1}^{2} P_{i,1975}\, Q_{i,1967}}{\sum_{i=1}^{2} P_{i,1967}\, Q_{i,1967}} \right]$$

$$= 100 \left[\frac{[(\$152.72)(1000) + (\$18.10)(100)]}{[(\$\ 84.12)(1000) + (\$15.20)(100)]} \right] = 180.4$$

All calculations using the Laspeyres method of index number construction use base year quantities for weighting numbers. In a similar fashion, the rest of the index numbers calculated by the Laspeyres method are:

Year	1967	1969	1971	1973	1975
Index	100.0	112.8	127.2	143.6	180.4

In 1874, H. Paasche presented another approach for constructing index numbers which uses current period quantities as weights. The formulation of the Paasche method is:

$$I_j = 100 \left[\frac{\sum_{i=1}^{m} P_{ij}\, Q_{ij}}{\sum_{i=1}^{m} P_{io}\, Q_{ij}} \right]$$

with the same meaning ascribed to the subscripts as previously defined. This index, like the Laspeyres index, is a ratio of weighted aggregates. But it relates the sum of current prices times current quantites to a hypothetical sum of base year prices times current quantities. Using the sample data for an illustration again:

$$I_{1975} = \left[\frac{\sum_{i=1}^{m} P_{i,1975}\, Q_{i,1975}}{\sum_{i=1}^{m} P_{i,1967}\, Q_{i,1975}} \right]$$

$$= \frac{[(\$152.72)(1015) + (\$18.10)(99)]}{[(\$\ 84.12)(1015) + (\$15.20)(99)]} = 180.5$$

In a similar fashion, the rest of the index numbers calculated by Paasche are:

Year	1967	1969	1971	1973	1975
Index	100	112.8	127.2	143.5	180.5

In the illustration, there was insufficient variation in the quantities purchased between the years to produce a significant change in the index numbers calculated. Nevertheless, when there are many items in the calculations, and the variation in the quantities being purchased is significant, the different approaches will produce different results.

The Paasche and Laspeyres weighted aggregative indexes are superior to the simple indices previously discussed. By the weighting approach, they are freed from problems of distortion associated with the size of the units for which the prices are quoted. The Laspeyres index is essentially the same one used by the Bureau of Labor Statistics in constructing their extensive series of price indexes. As can be seen from the differences in formulation, the data gathering task is simpler for the Laspeyres construction since only base year quantity data are needed. There is considerable argument in the literature as to which, if either, of these two approaches is better. More complicated approaches have been devised such as the "Fisher" ideal index to try to eliminate any bias introduced by selecting only base year or current year quantities for weighting. There does not seem to be any conclusion as to a best method for index number construction.

Weighted Average of Price Relatives Indexes

Another method of index number construction using a weighted average of price relatives is introduced here because it offers a practical means of combining previously constructed index numbers and specially constructed index numbers. This method is also well suited to using contractor accounting record data to construct the indexes. The formulation follows:

$$I_j = \sum_{i=1}^{m} I_{ij} \cdot W_{ij}$$ = weighted average of relatives index number for given time period j

where

$$I_{ij} = 100 \left[\frac{P_{ij}}{P_{io}}\right]$$ = price relative of the i^{th} commodity or service expressed in percentages; or a previously calculated index number from some source for the i^{th} commodity and j^{th} time period

i = subscript denoting the i^{th} commodity or service and ranging from 1 to m

j = subscript denoting the time period and ranging from the base period o to n

$$W_{ij} = \frac{P_{ij}\ Q_{ij}}{\sum_{i=1}^{m} P_{ij}\ Q_{ij}}$$ = the relative weight of the i^{th} commodity or service expressed as a relative of the total of all the commodities or services during the given time period

P_{io} = price (expressed in dollars per some unit of measure) of i^{th} commodity or service in base time period o

P_{ij} = price of i^{th} commodity or service in j^{th} given time period

Q_{ij} = quantity (unit of measure) of i^{th} commodity or service in j^{th} given time period

The approach is illustrated in Tables 9-1 through 9-5.

The illustration shows the computation of an employee benefits index number for the years 1967, 1969, and 1975 using a weighted average of price relatives, weighted by current expenditures.

One can observe that instead of constructing a price relative for a particular item or service, one could substitute a previously constructed index number for I_{ij} in the formulation. The ease of mixing previously constructed index numbers and specially constructed index numbers plus the advantage of using contractor cost data for weighting factors make it a practical method for the analyst to master.

Table 9-1

CONTRACTOR ACCOUNTING DATA

Year	1967	1969	1975
No. Indirect Employees	2000	2100	2150
Overhead Accounts			
1-Paid Absences	$1,262,000	$1,398,600	$1,388,900
2-Employee Insurance	552,000	598,500	604,150
3-Savings-Retirement	810,000	894,600	991,500
4-Education	18,000	21,000	21,500
Total Employee Benefits	$2,642,000	$2,912,100	$3,005,700

First, calculate prices from cost and volume data:

$$P_{ij} = \frac{\text{Amount paid in the year for service}}{\text{Quantity of service bought during the year}}$$

e.g., Price of Paid absences in 1967 = $\frac{\$1,262,000}{2000 \text{ person}}$ = \$631/person

Table 9-2

PRICES

Year	1967	1969	1975
Overhead Accounts			
1	$631	$666	$646
2	276	285	281
3	405	426	461
4	9	10	10

Second, calculate price relative (expressed as an index number) of average benefits per employee:

e.g., Price relative of paid absences in 1969 as compared to the base year 1967 is

$$100 \times \frac{P_{ij}}{P_{io}} = \frac{\$666}{\$631} = 105.5$$

Table 9-3

INDEX NUMBERS
(Base 1967) 100 x Price Relative

Year	1967	1969	1975
Overhead Accounts			
1	100.0	105.5	102.3
2	100.0	103.3	101.8
3	100.0	105.2	113.8
4	100.0	111.1	111.1

Third, calculate the relative weights of the benefits in each category for each year:

$$W_{ij} = \frac{P_{ij}\ Q_{ij}}{\sum_{i=1}^{m} P_{ij}\ Q_{ij}}$$

e.g., Calculate the relative weights of paid absences to total employee benefits for the year 1967:

$$\frac{\$1,262,000}{\$2,642,000} = .478$$

Table 9-4

RELATIVE WEIGHT OF BENEFITS

Year	1967	1969	1975
Overhead Accounts			
1	0.478	0.480	0.462
2	0.209	0.206	0.201
3	0.306	0.307	0.330
4	0.007	0.007	0.007
	1.000	1.000	1.000

Fourth, calculate index numbers for each year by summing the products of each account index number and its relative weight:

$$I_j = \sum_{i=1}^{m} I_{ij} \cdot W_{ij}$$

e.g., $I_{1968} = 100\ [(105.5 \times 0.480) + (103.3 \times 0.206) + (105.2 \times 0.307) + (111.1 \times 0.007)] = 105.0$

Table 9-5

EMPLOYEE BENEFITS INDEX NUMBERS

Year	1967	1969	1975
Index Nr.	100.0	105.0	106.1

It is a matter of judgment on the part of the index number constructor whether to use current year data or base year data to construct the weighting factors. In the above example, current year quantities are indicated. The weighting factors are constructed from accounting data of the type usually available from contractor records.

CHAPTER 10

TIME SERIES FORECASTING TECHNIQUES

This section introduces you to a limited number of basic, but nevertheless powerful, forecasting techniques for analyzing the trend in time series data.

The three basic groups of techniques for analyzing trend are: (1) line fitting techniques, (2) moving averages, and (3) exponential averages. All three of these groups of techniques are naive because they attempt to project the past into the future with minimal attention paid to actual cause and effect relationships. They are unsophisticated in that there are more powerful (and more complex) techniques available for analyzing time series data. Furthermore, this section does not pretend to cover even these naive and unsophisticated techniques in great depth. The use of line fitting techniques, moving averages, and exponential averages to analyze time series data goes well beyond the applications and examples presented here. Nevertheless, the techniques discussed are useful and relatively powerful in spite of, or perhaps because of, their simplicity. These techniques have been applied in the "real" world and have proven adequate enough to insure their continued use.

The techniques themselves are mathematically mechanical operations. Fed a certain amount of data, they will produce a number. It would be improper at this point, however, to label that number a forecast. All these forecasting techniques can do is process data into information. Forecasters forecast. Not until the forecaster applies his own knowledge, judgment, and good common sense to the problem can this information be developed into a forecast.

What are the characteristics of a good forecaster? First, he must have a good working knowledge of the tools of his trade. He must know the characteristics, limitations, strengths, and weaknesses of the analytical techniques available to him so that he can:

(1) Select the most appropriate technique for the situation at hand.
(2) Properly apply the technique.
(3) Interpret the information generated by application of the technique.

Second, he must be knowledgeable about the particular event or variable he is trying to forecast. The future does not just happen; it is the result of actions that have occurred in the past and/or are occurring in the present. If the forecaster attempts to forecast the future without having any conception of the causes underlying the past behavior of his data, he is trusting to blind luck.

Third, the forecaster must have good judgment and common sense. In spite of all the data and all the sophisticated techniques available to him,

/

in the final analysis, the forecast must be the product of his own thinking. If this were not so, forecasters could be replaced by bookkeepers and computer operators. Unfortunately, there is no such creature as a totally objective forecast. From the beginning to the end of the analysis, the forecaster must make decisions that influence the final outcome of his endeavors. If the results of his analysis are contrary to what the forecaster's experience and intuitive logic have led him to expect, then he must reconcile the differences. The forecaster must be certain that his forecast makes sense, that it is plausible, before he presents it to the decision maker.

Before we proceed any further in our discussion of forecasting, it is necessary to introduce some symbology. Throughout the text the symbology will consist of a combination of capital letters and numbers followed by a set of parenthesis containing small letters and numbers. The capital letters and numbers will identify the particular subject. The small letters and numbers contained in the parenthesis will be parameters that describe the subject. I realize that this is not clear, but a few examples should clarify what I am saying.

Let: Y = the variable we are going to forecast
F = our forecast of the variable Y
t = the time period from which we are making our forecast
i = the number of periods out from time period t that we are trying to forecast (forecast interval)
n = the number of observations used to make our forecast

Therefore: Y(t) = the value of variable Y in time period t
Y(t+i) = the value of variable Y in time period t+i
F(t,t+i) = our forecast of variable Y made in time period t for time period t+i

The capital letters Y and F tell us whether we are talking about the actual value of a variable or the forecasted value of a variable, and the small letters t and i are used to describe the particular Y or F we are talking about.

Since I am using parenthesis to contain descriptive information, I cannot use parenthesis for multiplication. In other words, Y(t) does not mean Y "times" t. Therefore, whenever I want to indicate that multiplication is to take place, I will use an asterisk "*". Therefore, Y(t+i) will mean the value of variable Y in time period t+i, and Y(t+i)*(t+i) means the value of variable Y in time period t+i times the value (t+i).

Using our new symbols, a forecasting exercise would go like this. We are in time period t. We know the value of our variable Y in the current time period; it is Y(t). We want to forecast the value of variable Y in some future time period, t+i. Our forecast, made in time period t, is F(t,t+i).

Next, we need to ask the question, what are the characteristics of a good forecast. One important characteristic is that the forecast be accurate. It would be the best of all possible worlds if the future turned out to be a perfect mirror image of our forecast. Using our symbology, perfect accuracy would exist when our forecast of variable Y made in time period t for time period t+i equaled the actual value of Y in time period t+i, or F(t,t+i) = Y(t+i). It is apparent from the equation that we cannot measure the accuracy of our forecast until we reach the time period we are trying to forecast. If our forecast was good, we can pat ourselves on the back and take our bows. If the forecast was bad, we can hide out until the repercussions of our error have dimmed in the boss's mind. The one thing we can't do is rectify the situation. Just like in football, no amount of Monday morning quarterbacking will change the final score.

Unfortunately, it is the forecaster's plight that the probability of F(t,t+i) = Y(t+i) is very close to zero. When it comes to absolute forecasts of the future, forecasters are almost always wrong. The goal in forecasting is not perfection. The perfect forecast is yet to be made. The goal in forecasting is to be able to predict the future with enough accuracy that the errors in your forecast are both consistent and small.

We can measure the error in our forecast by defining a new symbol, E, which stands for forecasting error.

$$E(F:t,t+i) = Y(t+i) - F(t,t+i) \qquad \text{(Equation 10-1)}$$

The error in our forecast is equal to the difference between what actually occurred, Y(t+i), and what we forecasted would occur, F(t,t+i).

For example, let us suppose that the following table represents our record of forecasts made in the past using a forecasting technique which we will refer to as F1.

Table 10-1

Y(t+i)	F1(t,t+i)	E(F1:t,t+i)
293	309	-16
312	322	-10
347	334	13
369	387	-18
384	367	17
420	406	14
441	454	-13
468	451	17
495	482	13
516	533	-17

The question we must now ask is how do we measure the accuracy of our forecast. To answer this question, we must examine the error term, E(F1).

One possible statistic we could use to describe the error term would be the range, i.e., the highest and lowest values for E(F1). In our example, the range of our error term is from -18 to +17. The smaller the range of values for the error terms, the more accurate our forecast. Suppose someone made a forecast and told you the range of their past forecasting errors was -$9,000,000 to +$8,000,000. What would your conclusion be regarding the accuracy of their forecasting technique? The question cannot be answered because you have no knowledge regarding the value of the variable being forecasted. If they were forecasting a value of $20,000,000 the range of their error term suggests a very inaccurate forecast. If their forecasted value was $2 billion, the range of the error term seems very small by comparison. It would seem worthwhile, therefore, to compare the range of our error terms to the values of the variable being forecasted.

One method would be to express the range of the error term as percentages of the mean of Y(t), or $\overline{Y}(t)$. In our example from Table 10-1, $\overline{Y}(t)$ = 4045/10 = 404.5. Therefore, the range of our error term expressed as a percent of $\overline{Y}(t)$ would be -4.45% to +4.20% of $\overline{Y}(t)$. This method will give you a fairly consistent measure of the magnitude of the error term when Y(t) fluctuates around $\overline{Y}(t)$ over time and the variation in E(t) around $\overline{E}(t)$ is relatively constant. A second method would be to convert each error term to a percent of the observation being forecasted and express the range of the error term as the smallest and largest percentage error. In our example from Table 10-1, the smallest error term is -5.46% and the largest error term is +4.43%. These error terms correspond to Y(t+i) = 293 and Y(t+i) = 384, respectively. Therefore, using this method, we would express the range in our error term as -5.48% to +4.43% of Y(t). This method will give you a fairly consistent measure of the magnitude of the error term when Y(t) is increasing (or decreasing) over time and the error term is increasing (or decreasing) over time at the same relative rate.

Another possible statistic would be the average value of the error terms $\overline{E}(F1) = \Sigma E(F1)/n$, where n equates to the number of error terms. However, when we add up the error terms in our example, $\Sigma(E(F1) = 0$ and $\overline{E}(F1) = 0$. We can see from our data that the negative errors and the positive errors are cancelling each other out. This could happen whether the error terms are large or small. Consequently, $\overline{E}(F1)$ does not seem to convey any significant information about the characteristics of the error team. One method for eliminating the effect of negative and positive signs is to examine the squared error terms. We can find the sum of the squared error terms and use this number to calculate a statistic that measured the dispersion of our forecasts around the actual observed values. In our example, $\Sigma E^2(F1)=2250$. Dividing this sum by the number of error terms and we get $\Sigma E^2(F1)/n = 2250/10 = 225$. This is the average of our squared error terms. Therefore, we must take the square root, $\sqrt{\Sigma E^2(F1)/n} = \sqrt{2250/10} = \sqrt{225} = 15$. This number we will call the standard error of the forecast.

$$S(F) = \sqrt{\frac{\Sigma E^2(F:t,t+i)}{n}} \qquad \text{(Equation 10-2)}$$

S(F) measures the dispersion of our forecasts around the actual values of Y. The smaller the standard error of the forecast, the more accurate our forecast.

Another method of counteracting the cancelling effect of positive and negative signs is to take the average of the absolute values of the error term. When you take the absolute value of a number, you treat the number as though it were positive regardless of its original sign. Going back to our example $\Sigma|E(F1)| = 148$. The average of the absolute values would be $\Sigma|E(F1)|/n = 148/10 = 14.8$. This number is used frequently to measure the accuracy of forecasts, and is called the mean absolute deviation or MAD, where

$$MAD(F) = \frac{\Sigma|E(F:t,t+i)|}{n} \qquad \text{(Equation 10-3)}$$

The MAD(F) is always smaller than the standard error of the forecast, and the smaller the MAD(F) the more accurate our forecast.

In statistics, the standard error (or standard deviation) was the measure of dispersion most used to describe the behavior of a data set. This was true because in statistics we dealt with probabilities, and our probabilities were determined by the standard error of our particular probability distribution.

In time series forecasting, however, we cannot do probabilistic forecasting (within the limits of this discussion). Our measure of dispersion is used to give us an indication of the expected accuracy of our forecast. Therefore, we should focus our attention on the measure of dispersion that has the greatest information value, i.e., we can understand what it is telling us.

In my opinion, the concept underlying the standard error is a nebulous concept at best. Just what does the average of the sum of the squared deviations of the error term around its mean measure? Does the concept become any more clear when you take the square root? Stripped of its probabilistic characteristics, the standard error as a measure of dispersion is very low in information content. It does not speak to me.

The MAD(F), by comparison, is informative. The fact that the MAD(F) cannot be associated with a particular probability distribution is irrelevant in time series forecasting. The MAD(F) does speak to me. It tells me, on the average, how much my forecasts have varied from the observations they were trying to predict.

Our three measures of forecasting accuracy, the range, the standard error of the forecast, and the MAD(F) do not always point to the same forecasting technique as being the most accurate. For example, suppose we had a second forecasting technique labeled F2, which provided the following results.

Table 10-2

Y(t+i)	F2(t,t+i)	E(F2:t,t+i)
293	278	15
312	320	-8
347	321	26
369	360	9
384	392	-8
420	441	-21
441	432	9
468	493	-25
495	503	-8
516	504	12

The range for E(F2) goes from -25 to +26. The $\Sigma E(F2) = 1$, so $\bar{E}(F2) = .1$. The $\Sigma E^2(F2) = 2465$, so $S(F2) = \sqrt{2465/10} = \sqrt{246.5} = 15.7$. The $\Sigma|E(F2)| = 141$, so $MAD(F2) = 141/10 = 14.1$.

To evaluate F(2) against F(1) we must compare their respective error terms. We do this be comparing the statistics which describe the behavior of the error terms.

Table 10-3

Statistics	F1	F2
Range	-18 to 17	-25 to 26
S(F)	15.0	15.7
MAD(F)	14.8	14.1

The range of values for the error term is smaller for F1 than F2 and so is the standard error of the forecast. F2, however, has a smaller mean absolute deviation. Since our previous discussion lead us to the conclusion that the MAD(F) is a more informative measure of forecasting accuracy, why not ignore the standard error of the forecast altogether? The answer is that there is a relationship between S(F) and the MAD(F) that reveals information about the behavior of the error term which is not communicated by the MAD(F) alone. The closer the absolute value of the error terms are to the MAD(F) the closer the S(F) will be to the MAD(F). Conversely, the further away the absolute value of the error terms are to the MAD(F), the further away S(F) will be from the MAD(F). In our example from Table 10-1, the absolute values of the error terms fluctuated around the MAD(F1) with little deviation. Five of the ten error terms are within the range of the MAD(F1) $\pm$ 3. For F1, the S(F1) and MAD(F1) are reasonably close together, 15 versus 14.8. In our example from Table 10-2, the absolute values of the error terms fluctuate widely around the MAD(F2). Only one of the ten error terms are within the range of MAD(F2) $\pm$ 1, and only two error terms fall within the range of the MAD(F2) $\pm$ 2. You have to extend the range to the MAD(F2) $\pm$ 7 to pick up eight of the ten error terms. For F2, the S(F1) and MAD(F2) are further apart, 15.7 versus 14.1. The S(F) and the MAD(F) are equal to each other only when the absolute value of the error term is a constant for every observation. The greater the variability in $|E(F)|$ around the MAD(F) the greater the difference between the S(F) and the MAD(F).

Therefore, we can expect that when the absolute values of the error term fall within a small range, the S(F) and the MAD(F) will be close together. Also, if S(F) and the MAD(F) are close together, we know that the forecasting technique under investigation produces error terms that are of a consistent size. We can also expect that when the absolute value of the error terms occupies a large range, S(F) and the MAD(F) will be further apart. Likewise, when S(F) and the MAD(F) are far apart, we know that the forecasting technique under investigation produces error terms that vary significantly in size.

The forecasting error for F1 is more consistent from period to period. You can expect your forecasts to be within plus or minus 15, give or take one or two. Relying on F2, however, you cannot predict whether your next forecast will be right to within plus or minus 25. Therefore, risk of error becomes more difficult to manage with F2 than with F1. However, F2 gave more accurate forecasts in seven out of ten times, while F1 was more accurate only three times out of ten times. So, using accuracy as our criteria, which is the better forecasting technique? It appears that there is no one unequivocal answer to this question. We must fall back on the old adage, it depends on the situation. If your situation is such that you cannot tolerate large forecasting errors because the cost of missing a forecast by a large amount is unacceptable, you would select F1 for your forecasts. If you are in a situation where you must reserve significant amounts of scarce resources to protect yourself against the cost of large forecasting errors, F1 would require a smaller reserve. If you are in a situation where the benefits derived from accurate forecasts exceed the costs of large forecasting errors, and you can afford the cost of an occasional forecasting error of large proportions, you would choose F2. If your success depends upon being right more often than you are wrong, you may be forced to choose F2 in spite of the risk of a large error.

As you can see, the selection of an appropriate forecasting technique depends upon three things in addition to the measured accuracy of the forecasting technique itself: (1) the cost to you of making a forecast that turns out to be very wrong, (2) the benefits derived from making a forecast that turns out to be very accurate, and (3) your ability to tolerate risk.

Another characteristic we would like our forecast to have is that it be responsive to changes in the environment. Almost all forecasts are extrapolations of the past into the future, from the known to the unknown. However, the past does not necessarily repeat itself in duplicate. Events change, and these changes will only become evident as they appear in our data base. Therefore, a forecasting technique would be responsive to changing conditions only if it was very sensitive to changing values reflected in the most recent observations.

However, there is a tradeoff. Events seldom occur in nice regular patterns. The variable we are trying to forecast is subject to random fluctuations, sometimes up and sometimes down, because of the combined effect of a multitude of causation variables that are individually trivial in

their impact. These random fluctuations will overlay the primary cause and effect relationships determining our forecasted variable. We do not want our forecast to change because of random fluctuations when the primary cause and effect relationships are stable. We do want our forecasts to respond to changes in the primary cause and effect relationships. However, if we design our forecasting technique to be sensitive to changes in the most recent data in order to accomplish our latter objective, our forecasts will also respond to random fluctuations as we collect new data. If we desensitize our forecasting technique to random fluctuations we lose the ability to detect changing conditions. How do you accomplish both objectives? The answer is, you don't.

As a forecaster you are confronted with the eternal dilemma of choosing between responsive and stable forecasts. Your choice will depend on the particular situation. You will want to use a forecasting technique that is responsive to changing conditions even at the expense of distortion caused by random fluctuation whenever: (1) the probability that conditions will change is high, (2) the cost of making a decision based on an erroneous forecast is high, (3) the benefit of being able to respond quickly to changed conditions is high, (4) the data is subject to relatively small random fluctuations, and (5) the cost of making decisions based on forecasts distorted by random fluctuations is low.

You will want to use a forecasting technique that produces relatively stable forecasts whenever: (1) the data is subject to large random fluctuations, (2) the cost of making decisions based on forecasts distorted by random fluctuations is high, (3) the probability that conditions will change is low, (4) the cost of making a decision based on erroneous forecasts is low, and (5) the benefit of being able to respond quickly to changed conditions is low. It is unfortunate that the stipulations mentioned above usually do not occur simultaneously in a way that makes the choice between stability and responsiveness an obvious one. There is never one right answer. You can only make reasonable choices based on judgment and common sense and then learn to live with the risk inherent in an uncertain future.

A third desired characteristic of a forecast is that it be reliable. Every forecasting technique is based upon its own set of assumptions which underly and, to a certain extent, determine the forecast. Reliability, as a forecast characteristic, depends upon the continued validity of these assumptions in the future. The assumption underlying the forecast must remain valid over time from the present to the period being forecasted before the forecast can be considered reliable. As the possibility increases that the assumptions upon which the forecast is based are losing their validity, the forecast itself becomes more and more unreliable.

The purpose of a forecast is to provide information for the decision making process. Therefore, the last characteristic we want our forecasting technique to possess is a favorable ratio of cost to benefits. The cost of making a forecast can be measured in three ways, money cost, time cost, and psychological cost.

The techniques discussed in this text are relatively cheap to make in terms of dollar cost. The greatest expense is probably occurred in the performance of the data gathering function. Once data is placed in a file accessible to the computer, programs written for the types of techniques with which we are concerned cost very little money to run. It is unlikely that cost would be a determining factor in your selection unless you move on to more sophisticated forecasting techniques.

Time may become a critical element in selecting a forecasting technique, even at our naive level. If a decision must be made tomorrow, a forecasting technique that will not produce a usable forecast in the amount of time available is worthless. Therefore, time constraints may restrict the choices you can make to respond to a particular problem.

Psychological costs are more nebulous and difficult to define. What is the cost to a decision maker when he is faced with making a decision based on information produced by a forecasting technique he does not understand and therefore, has little confidence in? A forecasting technique that can be understood by the decision maker because it falls within his realm of competence may be a better choice over a sophisticated but complicated forecasting technique completely beyond his comprehension. The role of the forecaster is to aid and assist the decision maker, not to just make forecasts. Consequently, the acceptance by the decision maker of the forecaster's work may be as critical as the forecast itself if the forecast is to be of any real value.

Good forecasts enhance the decision making process in two ways. First, a good forecast may provide some knowledge about a future that was heretofore unknown. This expands the information base upon which the decision maker must make his decision. By reducing the unknowns in the decision making process, the decision maker may be able to make more informed, objective, and therefore more reasonable decisions.

Second, a good forecast may confirm or support previously known information or reinforce intuitive knowledge. The forecast may not result in a different decision, but it will reduce the risk or uncertainty inherent in making the decision. The benefits are primarily psychological. Being able to make decisions in an environment of reduced risk or uncertainty alleviates much of the pressure (and fear) of both making and having to live with the consequences of a bad decision.

In conclusion, we want a forecasting technique that will deliver accurate forecasts, be responsive to changing conditions while immune to the effects of random fluctuations, and will deliver benefits to the decision maker in excess of its costs.

Anytime we have a variable for which we have measurements over time, time series analysis can be applied provided that two essential assumptions are found to be valid. Time series analysis attempts to define the path of a particular variable over time, a path that will be defined by some mathematical relationships which says $Y(t) = f(t)$.

The mathematical symbol "f", meaning "function of", implies a cause and effect relationship. To say that Y(t) is a function of time implies that time causes Y(t) to occur. For almost all time series data this is definitely not true. The variable Y(t) is almost never caused by time. However, if you believe that every effect has a cause, Y(t) must be caused by something. Furthermore, whatever is causing Y(t) to occur is not being accounted for in our analysis.

If you are doing an analysis and you want to omit a variable, there are two ways to do it. One method is to assume that the effect this one variable will have on the final outcome of your analysis is trivial or negligible, and therefore, adding the variable to your analysis only complicates matters without necessarily improving on the final product. In the case of time series analysis, we are talking about the actual causation variables, so this argument will not perform the deed for us.

The second method for omitting variables from an analysis is to assume (1) that the variable you wish to omit is not changing, and (2) that the effect this variable is having on your analysis is also unchanging. Both conditions are necessary. For example, suppose you are trying to forecast a variable Q which is caused by variable R along with other variables. If the value of R remains constant, and if the effect R has on Q also remains constant, you can omit R from your analysis.

It is this second method which we fall back on in time series analysis. However, we are looking at changes in Y(t) over time. If all of the causation variables do not change and the effect these variables are having on Y(t) do not change, Y(t) itself will not change. We know that Y(t) changes over time, so we must modify our assumption slightly. If the variables which are causing Y(t) to change continue to move along the same path in the future as they have in the past, and they continue to exert their effects on Y(t) with the same magnitude of force and direction as they have in the past, the causation variables can be omitted from our analysis. Only when both of these conditions are present can we base our forecast of Y(t) on the variable's past history alone.

An obvious, but critical, conclusion should come leaping into your mind at this point. YOU CANNOT USE TIME SERIES ANALYSIS TO FORECAST A VARIABLE WHEN YOU KNOW NOTHING ABOUT THE FORCES THAT ARE ACTUALLY CAUSING THAT VARIABLE TO CHANGE OVER TIME. The validity of the above two assumptions about the behavior of the causation variables is essential for a valid time series forecast. If either one of the two assumptions are violated, the time series forecasting techniques discussed in this course are not applicable.

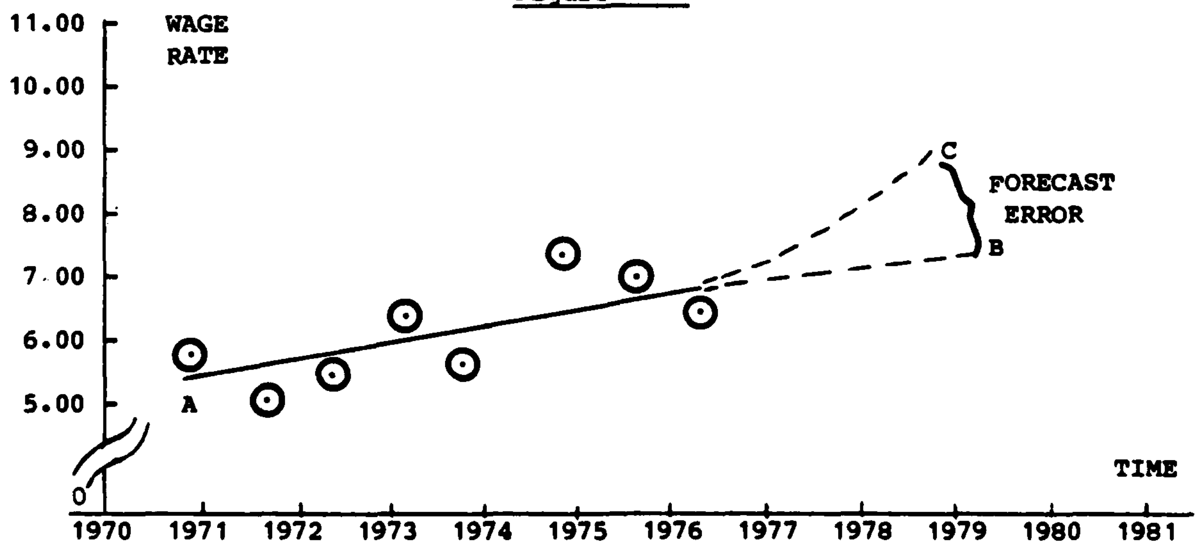
It must be equally obvious that to know whether or not the assumptions are valid, you must know something about the causation variables. As a minimum, you must be able to identify the major causation variables, and determine whether or not there is any basis for concluding that (1) they are not going to continue to move along their historical paths, and (2) they are not going to continue to exert their influence through historical cause and effect relationships.

Generally, it is not unrealistic to accept the validity of the two basic assumptions underlying time series forecasting unless there is evidence to the contrary. If you cannot find any plausible reason for believing that one or both of the assumptions are not valid, proceed with your forecast. However, if you want to have any confidence in your forecast, your investigation of the causation variables must be a serious one. If the assumptions are not valid, you can have no more faith in your forecast that you could have in a pure guess. The mathematical sophistication of the forecasting technique itself will be only illusionary; it will add nothing to the quality of your prediction.

It is worthwhile to note that the further out into the future we project the forecast, the more likely it will be that at least one of our assumptions will not hold true. If either one, or both, of the assumptions is not valid, the further into the future we try to predict, the more error we are likely to have in our forecast. Consequently, under most circumstances, forecasts based on time series analysis are limited in their ability to see very far into the future with any degree of reliability. The techniques discussed in this text are good for making short range forecasts, but their value for long range forecasts is limited by their underlying assumptions.

For example, suppose we had plotted wage rates against time (Figure 10-1). Our historical data goes from 1970 through 1977.

Figure 10-1



If we were to make a forecast from this data, we could easily assume from the plot-points alone that the wage rate is increasing along the straight line path AB. However, suppose that in 1975, the cost of living started to increase at a faster rate than it had in the past. To keep up with the cost of living, workers would demand higher and higher wage increases. In this case, our assumption that the causation variables must continue along their historical path has been violated. The cost of living is growing at a different rate than before. Therefore, wage rates may increase along the line AC. If we erroneously conclude that our assumptions are valid and forecast accordingly, you can see that the error in our forecast (the vertical difference between AB and AC) increases the further out into the future we try to predict. Furthermore, we have no way of measuring the potential magnitude of our forecasting error given that we make the wrong assumptions. Therefore, the further out into the future you need to forecast, the more thorough your analysis of the causation variables must be.

Let us suppose that we are in a situation where we can validly apply time series forecasting techniques to our data. Where do we start? In time series analysis we are looking for the path along which we expect our variable to move in the future. This path can be divided into four basic elements: trend, seasonal fluctuations, cyclical fluctuations, and random fluctuations. In this section we are going to ignore seasonal and cyclical fluctuations by assuming that these elements are not present in our data. Therefore, we are only going to be concerned with the analysis of trend and the presence of random fluctuations.

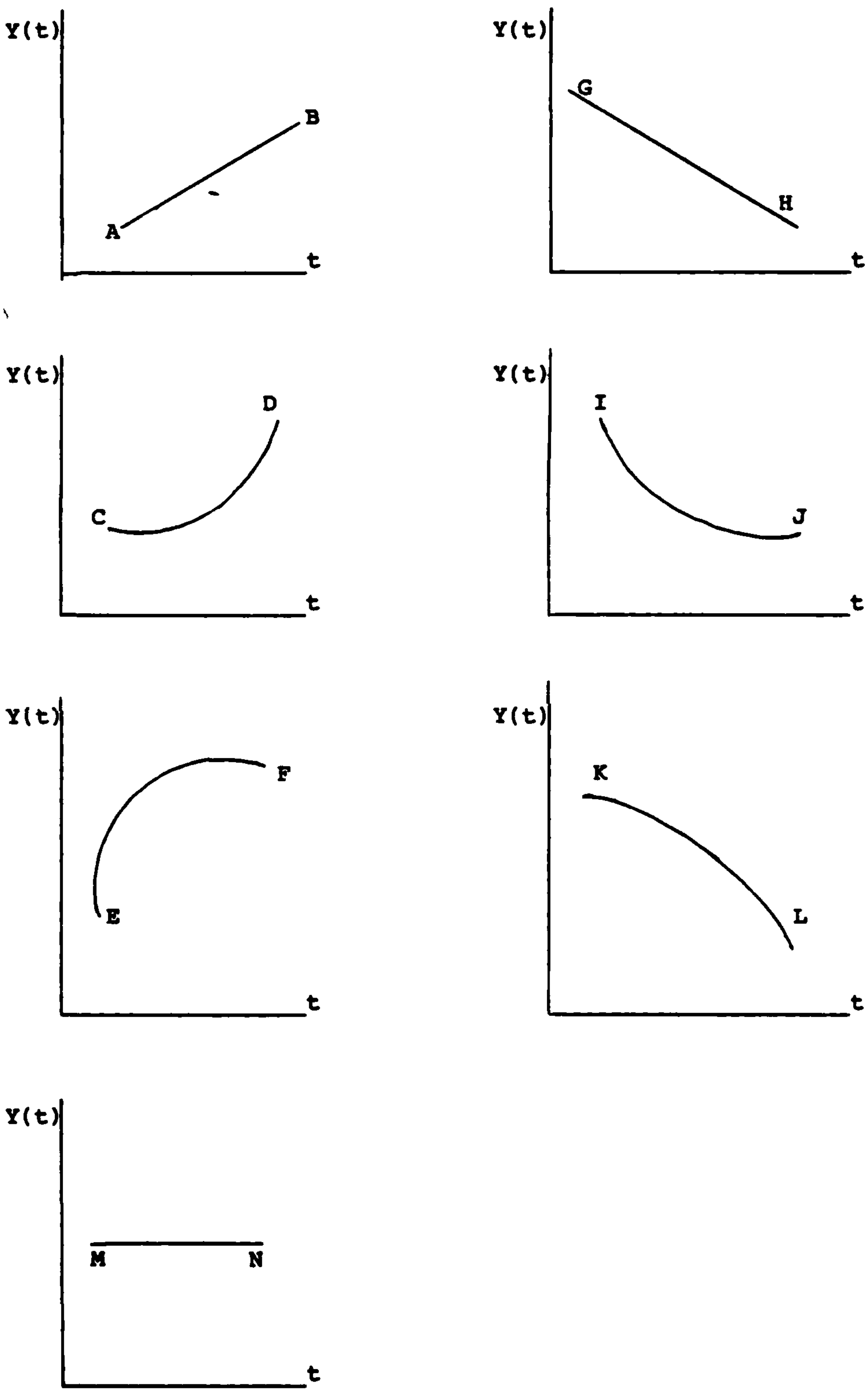
Trend is the primary direction of the path the variable will follow over time. In Figure 10-1, the lines AB and AC would represent two different forecasts of the trend in wage rates. There are seven basic trend lines that a time series variable can follow.

In Figure 10-2, lines AB, CD and EF all represent trends where Y(t) increases over time. The differences among the three lines reflect the differences in the rate at which Y(t) is increasing. Lines GH, IJ, and KL all represent trends where Y(t) decreases over time. Line MN represents a situation which we shall call zero trend. With zero trend, Y(t) fluctuates around a central value which remains constant over time.

Determining the location of a trend line requires three tasks. First, you must decide whether there is an increasing, decreasing, or zero trend. Second, if the trend line is increasing or decreasing, you must decide whether the trend is linear or nonlinear. Third, you must fit a trend line to the data in order to be able to extrapolate from the trend line into the future.

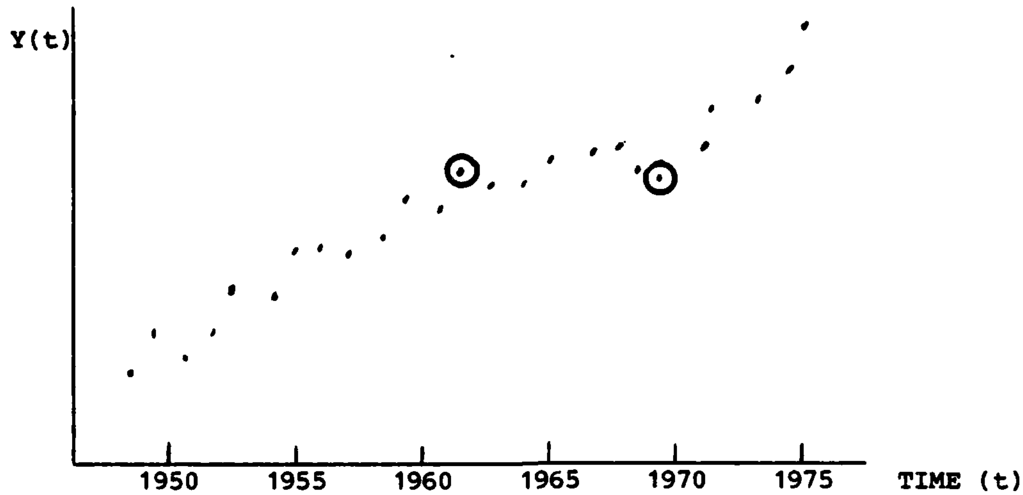
The first step in locating the trend line is to determine the number of data points you wish to include in your analysis. The trend line is a time path, and you want to find the trend line for your variable Y(t) in time period NOW. Therefore, you want to limit the observations in your data set to those data points that reflect Y(t)'s current trend line.

Figure 10-2



However, trends do change over time. In order to know how many observations you can include in your data set you must be able to determine when the last change in trend occurred. An excellent starting point would be to plot the data and take a look. Suppose your plot looks like Figure 10-3.

Figure 10-3



At what point in time did your trend line change it's course? You cannot get your answer from a graphic analysis. The best a graphic analysis can do is identify those points in time where possible changes in the trend line may have occurred. On our plot it appears that the trend line may have changed course in 1963, and again in 1970.

Having identified these two potential points in time, how do we determine if, in fact, a change in the trend line did occur. You must look to the cause and effect variables and their relationships to Y(t) to find the answer to your question. If the trend line did change, it changed for a reason. Either one or both of two events must have occurred. Either the cause and effect variables determining Y(t) shifted from their historical paths, or there has been a change in the historical cause and effect relationships, or both. In other words, you must test the two basic assumptions behind time series analysis and determine whether or not they remained valid during those points in time where you suspect a possible change in the trend line may have occurred.

Suppose your analysis revealed no basis for assuming a change in the trend line in 1970. However, there were changes in 1963 that support the hypothesis that 1963 was the beginning of a new trend. You have now answered your question about the number of observations to include in your data set for analysis. The current trend line has been continuing since 1963, the date of the last change in trend. Therefore, all data prior to 1963 must be discarded in your analysis. Pre-1963 data represents a different trend line and incorporation of that data in your analysis will only distort your results.

Having identified the observations that reflect the current trend, you must now decide whether or not you can reliably project that trend into the future. At this point a caveat is in order regarding time series analysis. Forecasts based upon time series analysis are totally incapable of predicting future changes in trend. They may react to actual changes in trend as new observations are accumulated that reflect the change, a characteristic that we have labeled responsiveness, but they have no predictive value. Therefore, every time series forecast is based on the assumption that there will be no future changes in the current trend.

To determine whether or not you can reliably project the current trend into the future as the basis for your forecast, you must again examine the basic assumptions underlying time series analysis, only this time you must speculate about the future. Will the assumptions remain valid from the present to the point in time that you are trying to forecast? And again you must rely upon the basic guidance that says if, after a thorough investigation, you fail to find evidence to the contrary accept the continued validity of the basic assumptions as reasonably probable, and use good judgment when incorporating the forecast into the decision making process.

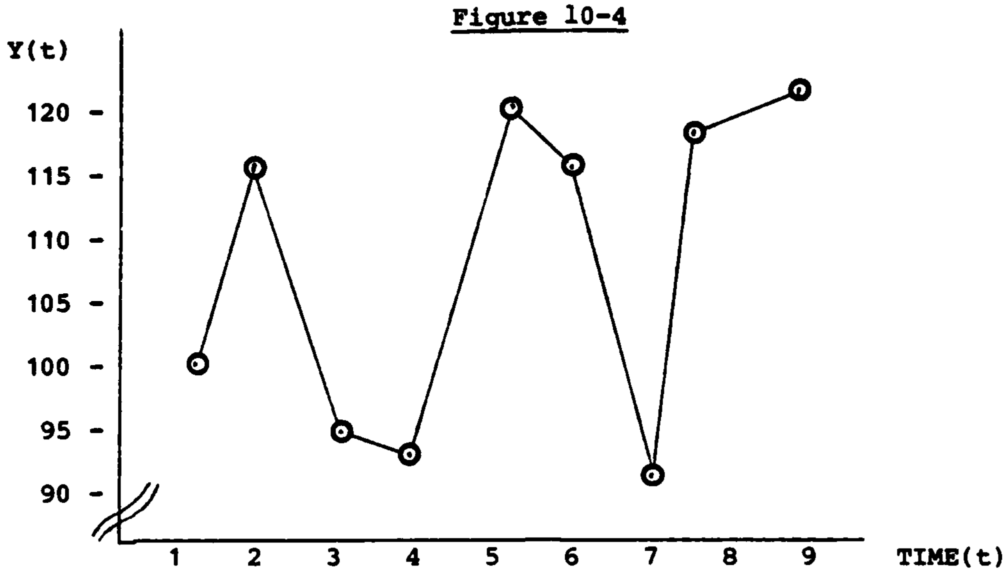
Once you have decided which observations to include in your analysis you are ready to begin your search for the trend line. Your first task is to determine whether there is an increasing, decreasing, or zero trend. Suppose you had the following data.

Table 10-4

t	Y(t)
1	100
2	113
3	97
4	95
5	118
6	114
7	94
8	115
9	119

A graph of our data would look as follows:

Figure 10-4



If our graph of the data revealed an obvious upward or downward trend, we would have answered our question and could then proceed to the next step of our analysis. However, for our data the trend, if it exists, is not obvious (at least not to me). Therefore, we need a statistical test that will give us a more conclusive answer when our graphic analysis fails us. The test we are going to use relies on a statistic called the Spearman Rank Correlation Coefficient, to which we will assign the symbol RS.

We have already listed our variable Y(t) in chronological order. By numbering time from 1 to n, we have "ranked" time from the past to the future. We must now rank our variable Y(t) from the smallest to the largest value. Our rank assignments will also go from 1 to n, with the smallest value of Y(t) being assigned the rank of "1", through the largest value of Y(t) which will receive a rank of "n". Our example would look as follows:

Table 10-5

Y(t)	Rank	Time
100	4	1
113	5	2
97	3	3
95	2	4
118	8	5
114	6	6
94	1	7
115	7	8
119	9	9

From the table we need to calculate the difference between the rank assignments for time and Y(t), which we shall call D(t). Finally, we must square each value of D(t) and find the sum of squares, or $\Sigma D^2(t)$. Once we have this number, we are ready to calculate RS. The formula is:

$$RS = 1 - 6*\Sigma D^2(t)/n^3-n \qquad \text{(Equation 10-4)}$$

For our example, the completed table would look as follows:

Table 10-6

Y(t)	Rank of Y(t)	Time	D(t)	$D^2(t)$
100	4	1	3	9
113	5	2	3	9
97	3	3	0	0
95	2	4	-2	4
118	8	5	3	9
114	6	6	0	0
94	1	7	-6	36
115	7	8	-1	1
119	9	9	0	0
				$\Sigma D^2(t) = 68$

Inserting $\Sigma D^2(t) = 68$ and $n = 9$ into our equation.

$$RS = 1 - 6*68/9^3 - 9 = 1 - 408/720 = 1 - .5667 = .4333$$

We now know that the Spearman Rank Correlation Coefficient for our data is .4333. Now what? To determine whether or not there is a trend, we must have a value to which we can compare our RS in order to draw some kind of conclusion. We will call this value RS(Critical). A table for RS(Critical) is provided. Your table for RS(Critical) is for a one-tailed test. Therefore, your hypothesis is that RS = 0, which means there is no trend. If RS > 0, your alternative hypothesis is that there is an upward, or a positive trend, and if RS < 0 your alternative hypothesis is that there is a downward, or negative trend.

In our example, RS = .4333 so we are testing the hypothesis that RS = 0 against the alternative hypothesis that RS > 0. The table for RS(Critical) is for three levels of significance, $\alpha = .10$, $\alpha = .05$ and $\alpha = .01$. Therefore, we must decide at what level of significance we want to test our hypothesis. Suppose we select $\alpha = .05$, which is equivalent to a 95% level of confidence. Then RS(Critical) = .5833. If RS < RS(Critical), accept your hypothesis. In our example, .4333 < .5833, so we must accept the hypothesis that RS = 0 and that there is no trend in the data. Note, the table for RS(Criteria), gives you values for the positive side of this probability distribution. However, RS can take on any value between -1 and +1. Therefore, when comparing RS to RS(Critical), the sign of RS should be ignored. Remember, however, that the sign of RS determines your alternative hypothesis.

Spearman's Rank Correlation Coefficient is used to test the hypothesis of no trend only under the condition that the trend, if one exists, is monotonic. In other words, the observations used to compute RS must meet the following conditions:

(1) There must be no cyclical effect
(2) There must be no seasonal effect
(3) There must be no changes in trend.

The trend line must be either continuously increasing, continuously decreasing, or constant.

Because of the effect of random fluctuations in the data, the RS test may lead you to the wrong conclusion regarding the acceptance or rejection of your hypothesis. The probability of this occurring is measured by the level of confidence you select for testing your hypothesis. However, if the three conditions specified above for the observations are not met, you might as well flip a coin to decide whether or not there is a trend. Your chances of being right are almost as good.

If there is no trend in the data our variable must be fluctuating around a horizontal line with a slope of zero. In other words, you trend line is Y(t) = a. However, if we determine that there is a trend in our data, we must attempt to locate the trend line in order to project future values for our variable. To do this, we need to find the equation that relates Y(t) to time, t. We have a technique available to use to find Y(t) = f(t) when f(t) = a + b*t, a straight line. The technique is called least-squares-best-fit, hereafter referred to as LSBF.

The equations for finding the parameters a and b using the LSBF technique when time is the ingredient variables are:

$$\Sigma Y(t) = n*a + b*\Sigma t$$
$$\Sigma[Y(t)*t] = a*\Sigma t + b*\Sigma t^2$$

Suppose we had the following data:

Table 10-7

Month(1976)	Y(t)
Jan	112
Feb	127
Mar	147
Apr	166
May	177
Jun	194
Jul	217
Aug	232
Sep	245
Oct	268
Nov	281
Dec	298

Time is represented by months of the year, which makes it extremely difficult to find $\Sigma[Y(t)*t]$, Σt, and Σt^2. Therefore, we need to find a scheme for quantifying time. In this section we will adopt the following rules. The oldest observation in our data set will be given a value of one, and times periods will be numbered sequential in ascending order as you move toward the present. This is not the only scheme available for quantifying time but it is the only scheme that will be applied in this section. Our quantification of time would then look like Table 10-8.

Table 10-8

Y(t)	(t)	Y(t)*t	t^2
112	1	112	1
127	2	254	4
147	3	441	9
166	4	664	16
177	5	885	25
194	6	1164	36
217	7	1519	49
232	8	1856	64
245	9	2205	81
268	10	2680	100
281	11	3091	121
298	12	3576	144
2464	78	18447	650

Substituting $\Sigma Y = 2464$, $\Sigma t = 78$, $\Sigma t^2 = 650$, $\Sigma[Y(t)*t] = 18447$, and $n = 12$ into our LSBF equations gives us:

$$2464 = 12a + 78b$$
$$18447 = 78a + 650b$$

Solving the above two equations for a and b we find that a=94.8333 and b=17. We shall use the symbol YT to represent our trend line, followed by the parenthesis (t). The symbol "t" represents three things. First, it is the most recent observation in our data set. Second, since we have numbered our observations sequentially starting with one, t also equates to the number of observations in the data set. Third, t represents the time period from which all our forecasts will be made. Usually, but not always, t will be the current time period. In our example, YT(12)=94.8333+17*t.

If you remember our earlier symbology, F(t,t+i) was a forecast made in time period t for the time period t+i. To designate a forecast based on a projection of the trend line, I am going to use the symbol FT followed by the parenthesis (t,t+i).

In our example, the trend forecast would be FT(12,12+i) = 94.8333 + 17*(12+i). Using this equation, let us forecast the variable Y for June 1977, supposing we are now in December 1976. For December 1976, t = 12. June is six periods into the future. Therefore, i = 6 and t+i = 18. Our forecast would be: FT(12,18) = 94.8333 + 17*18 = 94.8333 + 306 = 400.8333 or 401.

What would our trend forecast be if our test for trend indicates there is no trend in the data? To answer this question, look at the LSBF equations themselves. No trend means that the slope of the trend line is zero, or b = 0. The first equation is $\Sigma Y(t) = n*a + b*\Sigma t$. If b = 0, then $\Sigma Y(t) = n*a$ or $a = \Sigma Y(t)/n = \overline{Y}(t)$. Therefore, our forecast $FT(t,t+i) = \overline{Y}(t)$ would never change as long as the assumptions of our time series analysis hold true.

To determine whether or not a trend is nonlinear, we must again analyze the error term E(t) that is obtained from our LSBF linear trend equation. Take the following example:

Table 10-9

Y(t)	Month(1977)	t
139	Mar	1
144	Apr	2
140	May	3
151	Jun	4
165	Jul	5
188	Aug	6
211	Sep	7
246	Oct	8
281	Nov	9

The first step we would take is to test to see if there is a trend by calculating RS.

Table 10-10

Y(t)	Rank	t	d(t)	$d^2(t)$
139	1	1	0	0
144	2	2	1	1
140	3	3	-1	1
151	4	4	0	0
165	5	5	0	0
188	6	6	0	0
211	7	7	0	0
246	8	8	0	0
281	9	9	0	0
				$\Sigma d^2(t) = 2$

$RS = 1 - (6*2/(9)^3-9) = 1 -(12/720) = 1 - .01667 = .9833$

For $n = 9$ and $\alpha = .01$, RS(Critical) = .7667

RS = .9833 > RS(Critical) = .7667, so we reject the hypothesis that RS = 0 and accept the alternative hypothesis that RS > 0. We are certain at the 99% level of confidence that there is an upward trend.

Our next step would be to measure the linear trend by using the LSBF equations.

Table 10-11

Y(t)	(t)	t^2	Y(t)*t
139	1	1	139
144	2	4	288
140	3	9	420
151	4	16	604
165	5	25	825
188	6	36	1128
211	7	49	1477
246	8	64	1968
281	9	81	2529
1665	45	285	9378

$$\text{LSBF:} \quad \Sigma Y(t) = n*a + b*\Sigma t$$
$$\Sigma[Y(t)*t] = a*\Sigma t + b*\Sigma t^2$$

$$\text{or:} \quad 1665 = 9a + 45b$$
$$9378 = 45a + 285b$$

Solving these two equations, we find that a = 97.25 and b = 17.55. Therefore, YT(9) = 97.25 + 17.55*t. We next calculate the error term, E(YT:9) = Y(t) - YT(9).

Table 10-12

Y(t)	YT(Mar 77,9)	E(YT: Mar 77,9)	Sign + or -
139	114.8	24.2	+
144	132.35	11.65	+
140	149.9	-9.9	-
151	167.45	-16.45	-
165	185.00	-20.00	-
188	202.55	-14.55	-
211	220.1	-9.1	-
246	237.65	8.35	+
281	255.2	25.8	+

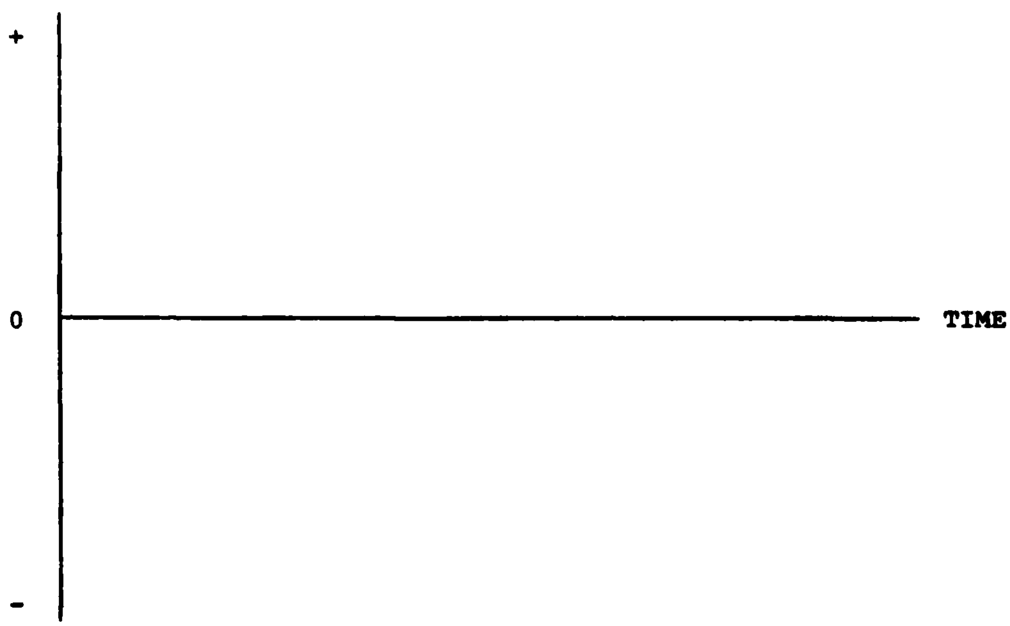
There are two methods for determining if the trend line is nonlinear. The first method is to look at the signs (+ or -) associated with the error term E(YT). If the signs are positive for the end values and negative for

the middle values, or negative for the end values and positive for the middle values, there is an indication that you have a nonlinear trend. If the signs appear to be "randomly" distributed, changing from positive to negative and back and forth, your trend is probably linear. In our example the signs associated with the end values are positive and the signs associated with the middle values are negative, indicating a nonlinear trend.

The second method is to plot the error term against time. Since the error term takes on both positive and negative values, your graph should be set up in the following manner.

EY(t)

<u>Figure 10-5</u>



If the chronological plot of E(YT) looks like a "⌣" or a "⌢" the trend line is probably nonlinear. If a plot of e(t) crosses the t-axis several times so the plot looks like random fluctuations, the trend is probably linear. In our example, the error terms plot line a "⌣".

The following tables give the relationship between the error terms and the nonlinear form of the trend line. The first column specifies whether the slope of the linear trend line projected through the data using LSBF is positive or negative. The second column indicates whether the signs of the error terms are positive at the ends and negative in the middle (+,-,+) or negative at the ends and positive in the middle (-,+,-). The third column indicates whether the error terms e(YT) look like a "⌣" or like a "⌢" when plotted against time. The last column indicates how the slope of the nonlinear trend is behaving.

Table 10-13

"b"	Sign of "E"	Plot of "E"	Nonlinear trend (Figure 10-2)
Positive	+,-,+		Trend line C-D
	-,+,-		Trend line E-F
Negative	+,-,+		Trend line I-J
	-,+,-		Trend line K-L

I am not going to attempt to provide you with the techniques necessary to determine the equation for a nonlinear trend using LSBF. Therefore, if you have a trend line which is nonlinear, I give you two choices. First, you can draw a free-hand line through the data, and attempt, based on a visual analysis, to graphically project the line to where you think it appears to be going. Second, you can select the most recent data points and, using LSBF, fit a straight line to the data and project. If you do not project too far out into the future, and the curvature of your nonlinear trend line is relatively flat, the error in your projection should not be excessive.

The LSBF line is not very sensitive to the addition of new observations. This lack of sensitivity to new observations has a twofold effect. First, our LSBF line will respond very slowly to changes in trend. Second, our LSBF line will remain relatively unaffected by random fluctuations in the data. If we find ourselves in a situation where stability in our forecast is more important than an ability to respond to changes in trend, a LSBF trend line is a reasonable technique upon which to base our forecast. However, if the reverse is true and the ability to respond to changes in trend gains greater relative importance, a LSBF trend line will not meet our needs. What to do? Eliminating older observations as we add new observations increases the sensitivity of our forecasting technique to new observations which, in turn, results in greater sensitivity to changes in trend. This is the basic idea behind the family of forecasting techniques based on moving averages. However, before we begin, we need to acquaint ourselves with some new symbology.

M1(n,t) is a single n-period moving average calculated in time period t.

M2(n,t) is a double n-period moving average calculated in time period t.

FM1(n,t,t+i) is a single n-period moving average forecast made in time period t for time period t+i.

FM2(n,t,t+i) is a double n-period moving average forecast made in time period t for time period t+i.

A single n-period moving average for time period t is the average value of the last n observations of a data set that ends at time period t. As you can see, verbally describing a single n-period moving average is very much like describing a spiral staircase without using your hands. Perhaps we can gain better insight into M1(n,t) by looking at the way a single n-period moving average is calculated. The formulas for M1(n,t) is:

$$M1(n,t) = \frac{Y(t) + Y(t-1) + Y(t-2) + . . . + Y(t-n+1)}{n} \qquad \text{(Equation 10-5)}$$

The summation in the numerator starts with the observation that corresponds to the time period for which we are calculating the moving average. There are also n observations in the numerator. In our symbology M1(n,t), t determines the first observation to be included in the moving average, and n determines the number of observations to be included. For example, suppose we were to analyze the following data:

Table 10-14

t	Y(t)
1	74
2	94
3	84
4	71
5	88
6	81
7	77
8	91
9	87
10	80

Before calculating M1(n,t) we must decide how many observations we want to include in our moving average. In fancier language, n is a decision parameter that must be selected by the forecaster based on factors that will be discussed later. For the purpose of our examples, let n = 3.

We cannot calculate M1(3,1) or M1(3,2) because we do not have a sufficient number of data points. The first n - period moving average we can calculate is M1(3,3).

Using equation 10-5 $M1(3,3) = \frac{Y(3)+Y(2)+Y(1)}{3} = \frac{84 + 94 + 74}{3} = \frac{252}{3} = 84$

To calculate M1(3,4) we drop the oldest observation in the numerator, Y(1) = 74, and add the newest observation, Y(4) = 71. Using equation 3-1,

$$M1(3,4) = \frac{Y(4)+Y(3)+Y(2)}{3} = \frac{88 + 71 + 84}{3} = \frac{243}{3} = 83$$

and

$$M1(3,5) = \frac{Y(5)+Y(4)+Y(3)}{3} = \frac{88 + 71 + 84}{3} = \frac{243}{3} = 81$$

I will leave it to you to calculate M1(3,t) for t = 6 to t = 10. Your final results should look like Table 10-15.

Table 10-15

t	Y(t)	M1(3,7)
1	74	-
2	94	-
3	84	84
4	71	83
5	88	81
6	81	80
7	77	82
8	91	83
9	87	85
10	80	86

The formula for a single n-period moving average can be modified slightly into a format that makes it more workable when n is large. We could write M1(3,5) as follows.

$$M1(3,5) = \frac{Y(5)+Y(4)+Y(3)}{3} = \frac{Y(5)}{3} + \frac{Y(4)}{3} + \frac{Y(3)}{3}$$

If we both add and subtract $\frac{Y(2)}{3}$ from the above equation, nothing changes. The two terms cancel each other out. The equation would then read:

$$M1(3,5) = \frac{Y(5)}{3} + \frac{Y(4)}{3} + \frac{Y(3)}{3} + \frac{Y(2)}{3} - \frac{Y(2)}{3}$$

However, the middle three terms represent the 3-period moving average for time period 4, M1(3,4). We can rewrite the equation:

$$M1(3,5) = \frac{Y(5)}{3} + M1(3,4) - \frac{Y(2)}{3} = M1(3,4) + \frac{Y(5)-Y(2)}{3}$$

We can calculate M1(3,5) by adding to M1(3,4) one third of the difference between the new added observation, Y(5), and the old deleted observation, Y(2).

In general,

$$M(n,t) = M(n,t-1) + \frac{Y(t) - Y(t-n)}{n} \qquad \text{Equation 10-6)}$$

In our example we calculated M1(3,3) = 84. Using equation 10-6.

$$M1(3,4) = M1(3,3) + \frac{Y(4) - Y(1)}{3} = 84 + \frac{71 - 74}{3} = 84 - 1 = 83 \text{ and}$$

$$M\ (3,5) = M1(3,4) + \frac{Y(5)-Y(2)}{3} = 83 + \frac{88 - 94}{3} = 83 - 1 = 81$$

When n is small it makes little difference whether you use equation 10-5 or equation 10-6 to calculate your moving average. However, as n gets large, use of equation 10-6 can save you a tremendous amount of time and effort. If you don't believe me, calculate M1(6,t) for t = 6 to t = 10 from the date in Table 10-14 using both equations. You will become a believer.

We now know how to calculate a single n-period moving average for time period t. However, suppose we are in time period t and we want to forecast time period t+i. What do we do? If you are going to forecast using a single n-period moving average, the relationship between the forecast and the moving average is a simple one.

$$FMI(n,t,t+i) = M1(n,t) \qquad \text{(Equation 10-7)}$$

Our single moving average forecast for time period t+i is nothing more than our single moving average for time period t. Going back to Table 10-15, suppose we wanted to make a forecast for two periods into the future. Then our forecast interval, i = 2. Our forecasts FMI(3,t,t+2) and the error in our forecasts EFMI:3,t,t+2) are shown in Table 10-16.

Table 10-16

t	Y(t)	M1(3,t)	FMI(3,t,t+2)	EFMI: 3,t,t+2)
1	74	-	-	-
2	94	-	-	-
3	84	84	-	-
4	71	83	-	-
5	88	81	84	4
6	81	80	83	-2
7	77	82	81	-4
8	91	83	80	11
9	87	85	82	5
10	80	86	83	-3

The sign of our error terms appears to be random. Our statistics for measuring forecasting accuracy would be:

$$MAD = \frac{|\Sigma EFM1|}{n} = \frac{29}{6} = 4.833$$

$$S = \sqrt{\frac{\Sigma (EFM1)^2}{n}} = \sqrt{\frac{191}{6}} = \sqrt{31.8333} = 5.642$$

When calculating our statistics we let n = 6 because there are only six periods for which we have forecasts even though ten observations were used in making our calculations.

Is there a trend in our data? To answer this question we must calculate the Spearman Rank Correlation Coefficient.

Table 10-17

t	Y(t)	Rank Y(t)	D	D^2
1	74	2	-1	1
2	94	10	-8	64
3	84	6	-3	9
4	71	1	3	9
5	88	8	-3	9
6	81	5	1	1
7	77	3	4	16
8	91	9	-1	1
9	87	7	2	4
10	80	4	6	36
				150

$$RS = 1 - \frac{6*150}{10^3-10} = 1 - \frac{900}{990} = 1 - .9091 = .0909$$

RS(Critical) for n=10 and an $\alpha = .05$ is .5515. Since RS < RS(Critical) we can accept the hypothesis that RS = 0 and be 95% confident of having made the right choice. We are 95% certain that there is no trend in the data.

Would it have made any difference in our forecasts if there had been a trend in the data? To answer this question let us look at another set of data.

Table 10-18

t	Y(t)
1	111
2	118
3	122
4	120
5	124
6	128
7	123
8	130
9	134

For the data in Table 10-18, RS = .9333. RS(Critical) for n = 9 and $\alpha = .01$ is .7667. Since RS is positive and greater than RS(Critical), we reject the hypothesis that RS = 0 and accept the alternative hypothesis that RS > 0. We are 99% certain that there is an increasing trend in the data.

Let us calculate a 3-period single moving average forecast for two periods into the future.

Table 10-19

t	Y(t)	M1(3,t)	FM1(3,t,t+2)	E(FM1: 3,t,t+2)
1	111	-	-	-
2	118	-	-	-
3	122	117	-	-
4	120	120	-	-
5	124	122	117	7
6	128	124	120	8
7	123	125	122	1
8	130	127	124	6
9	134	129	125	9

The error term for our forecast is always positive. This means our forecasts are consistently too low. What is the source of this bias? Because of the upward trend in the data the observations used to calculate the single moving average are usually smaller than the observations we expect to see in the future. Therefore, when the moving average is used as a forecast of the future, we would expect our forecast to usually be too low. If our trend line was decreasing over time, our single moving average forecast would usually be too high, and our terms would be negative.

The magnitude of the bias in our forecasts depends on n, i, and the slope of the trend line. The steeper the slope of the trend line, the greater the bias in your forecast. The greater the number of observations in your single moving average calculation, the greater the bias in your forecast. The greater the number of periods out into the future you are trying to forecast, the greater the bias in your forecasts. Conversely, the flatter the slope of the trend line, the smaller the value for n, and the smaller the value for i, the smaller the bias in your forecast. However, if you have a trend in your data, a single moving average forecast will always produce biased forecasts.

If you have a trend in your data you may still be able to forecast with a single moving average provided that the amount of the bias is relatively consistent. In other words, the bias in the forecast must be relative stable over time. The forecast error term is a combination of the amount of bias in the forecast and random fluctuations in the data. If there is no bias in the forecast, the error term measures only the random fluctuations, and the sum of the error terms should equal zero. However, if there is bias in the forecast the sum of the error terms should be positive or negative depending on whether the bias is positive or negative. Our best estimate of the average value of our bias, therefore, is the average value of our error terms, or $\frac{\Sigma E}{n} = \bar{E}$. Once we have a measure of our forecast bias we can adjust our forecast by the amount of the bias.

For example, using the data from Table 10-19, $\bar{E} = \frac{31}{5} = 6.2$. Therefore, our forecast adjusted for bias would be FM1(3,t,t+2) + 6.2. Our forecasts

would then look like Table 10-20.

Table 10-20

t	Y(t)	FM1(3,t,t+2)+6.2	E(FM1: 3,t,t+2)-6.2
1	111	-	
2	118	-	
3	122	-	
4	120		
5	124	123.2	.8
6	128	126.2	1.8
7	123	128.2	-5.2
8	130	130.2	-.2
9	134	131.2	3.8

Note that by adjusting our single moving average forecast by $\bar{E}$, the sum of the resulting error terms will always add to zero, a condition that implies no forecasting bias.

Since we cannot rely on our single moving average forecasts for all cases when there is a trend in the data, what are we to do? If the trend is linear you can still use moving averages for forecasting, but instead of using single moving averages you must use double moving averages. What is a double moving average? A double moving average is a moving average of single moving averages. Remember the formula for a single moving average.

$$M1(n,t) = \frac{Y(t) + Y(t-1) + Y(t-2) + \quad + Y(t-n+1)}{n} \qquad \text{(Equation 10-5)}$$

If you were to substitute single moving averages in place of individual observations in the above equation, you would have a double moving average.

$$M2(n,t) = \frac{M1(n,t) + M1(n,t-1) + M1(M,t-2) +\ldots+ M1(n,t-n+1)}{n} \qquad \text{(Equation 10-8)}$$

To see how we would calculate double moving averages, let us look again at the data used to calculate the single moving averages in Table 10-19.

Table 10-21

t	Y(t)	M1(3,t)
1	111	-
2	118	-
3	122	117
4	120	120
5	124	122
6	128	124
7	123	125
8	130	127
9	134	129

Our single moving average is a 3-period moving average, so our double moving average must also be a 3-period moving average. M2(3,5) is the first double moving average we can calculate, since time period 5 is the first time period for which we have three single moving averages. Using Equation 10-8.

$$M2(3,5) = \frac{M1(3,5) + M1(3,4) + M1(3,3)}{3} = \frac{122 + 120 + 117}{3} = \frac{359}{3} = 119.667$$

To calculate M1(3,6) we need to change the summation in the numerator by deleting M1(3,3) and adding M1(3,6). The equation for M2(3,6) then becomes:

$$M2(3,6) = \frac{M1(3,6) + M1(3,5) + M1(3,4)}{3} = \frac{124 + 122 + 120}{3} = \frac{366}{3} = 122$$

I will leave it to you to calculate M2(3,t) for t = 7 to t = 10. Your results should look like Table 10-22.

Table 10-22

t	Y(t)	M1(3,t)	M2(3,t)
1	111	-	-
2	118	-	-
3	122	117	-
4	120	120	-
5	124	122	119.667
6	128	124	122.000
7	123	125	123.667
8	130	127	125.333
9	134	129	127.000

Equation 10-5 can also be used to calculate double moving averages by changing the individual observations to single moving averages.

$$M2(n,t) = M2(n,t-1) + \frac{M1(n,t) - M1(n,t-n)}{n} \qquad \text{(Equation 10-9)}$$

For example, the double moving averages in Table 10-21 could have been calculated as follows:

$$M2(3,6) = 119.667 + \frac{124-117}{3} = 119.667 + 2.333 = 122$$

and

$$M2(3,7) = 122 + \frac{125-120}{3} = 122 + 1.667 = 123.667$$

Again, for a large n, equation 10-9 can save you substantial computational time compared to equation 10-8.

Two other observations are worth noting. First, it takes more data points to calculate a double moving average than it does to calculate a single moving average. A n-period single moving average requires a minimum of n observations, while a n-period double moving average requires a minimum of 2*n-1 observations.

Second, when there is a trend in the data the double moving average will lag behind the trend line even more than the single moving average. For example, you can see in Table 10-22 that when there is an increasing trend, M1(n,t) lags behind the trend. Furthermore, M1(n,t) becomes smaller as you move backward in time. When calculating M2(n,t) the numerator would consist of the sum of the most recent single moving average and n-1 earlier single moving averages smaller in value. Consequently, the average of these single moving averages, M2(n,t), would be smaller than the most recent single moving average, M1(n,t). When the trend is increasing $M2(n,t) < M1(n,t)$. Conversely, when the trend is decreasing $M2(n,t) > M1(n,t)$. When would M2(n,t) = M1(n,t)? When there is no trend and M1(n,t) and M2(n,t) are both measuring only the random fluctuation in the data.

When we based our forecast on a single moving average we used the relationship FMI(n,t,t+i) = M1(n,t). However, we saw that single moving averages gave us biased forecasts when our data contained a linear trend.

The equation for a straight line is Y = a+b*x. Since we are trying to forecast a linear trend, it would be reasonable to assume that the relationship between our double moving average forecast and our double moving average is also linear. In fact, a double moving average forecast depends on both the double moving average and the single moving average. The relationship is:

$$FM2(n,t,t+i) = a(t)+b(t)*i \qquad \text{(Equation 10-10)}$$

where

$$a(t) = 2*M1(n,t) - M2(n,t) \qquad \text{(Equation 10-11)}$$

$$a(t) = \frac{2}{n-1}\left[M1(n,t) - M2(n,t)\right] \qquad \text{(Equation 10-12)}$$

The double moving average forecast is a linear trend line where the parameters a(t) and b(t) are determined by the values of M1(n,t) and M2(n,t).

Using the values calculated in Table 10-22, our forecasting equation for time period t=5 would be:

$$a(t) = 2*M1(3,5) - M2(3,5) = 2*122 - 199.667 = 124.333$$

$$b(t) = \frac{2}{n-1}\left[M1(3,5) - M2(3,5)\right] = \frac{2}{3-1}\left[122 - 199.667\right] = 2.333$$

$$FM2(3,5,5+i) = 124.333 + 2.333*i$$

Using the forecasting equation, our forecast for time period 9, 4 time periods into the future, would be FM2(3,5,9) = 124.333 + 2.333*4 = 133.665. Our forecasting equations for time period t = 6 would be:

$$a(t) = 2*124 - 122 = 126$$

$$b(t) = \frac{2}{3-1} * [124 - 122] = 2$$

$$FM2(3,6,6+i) = 126+2*i$$

Our forecast of time period t = 9 from time period t = 6 would be FM2(3,6,9) = 126+2*3 = 132.

As we move closer and closer to time period t = 9 our forecast will continually change as we adjust our estimates for the linear parameters a and b for the new observations to our data.

What would happen if we were to base our forecasts on double moving averages when the data actually contains no trend? To answer this question, let us look at the equations for a(M2) and b(M2).

$$a(t) = 2*M1(n,t) - M2(n,t) \qquad \text{(Equation 10-11)}$$

$$b(t) = \frac{2}{n-1} *[M1(n,t) - M2(n,t)] \qquad \text{(Equation 10-12)}$$

If there is no trend in the data, M1(n,t) and M2(n,t) should theoretically be equal. In actual fact they probably will not be exactly equal because of the effects of random fluctuations in the data. Assuming that M1(n,t) does equal M2(n,t) what happens to Equation 10-11 and 10-12. Equation 10-12, the equation for b(t), consists of a multiplication of two numbers, one of which is the difference between M1(n,t) and M2(n,t). If M1(n,t) = M2(n,t), then M1(n,t) - M2(n,t) = 0 and Equation 10-12 will equal zero. The trend line will have no slope, which is consistent with our initial statement that the data contained no trend. Equation 10-11, the equation for a(t), consists of the difference between twice M1(n,t) and M2(n,t). If M1(n,t) = M2(n,t), then substituting M1(n,t) for M2(n,t) in the equation results in a(t) = M1(n,t). Since b(t) = 0, the equation for a double moving average forecast becomes:

$$FM2(n,t,t+i) = a(t) + b(t) *i = a(t) = M1(n,t) = FM1(n,t,t+i)$$

When there is no trend in the data, our double moving average forecast degenerates into a single moving average forecast.

If there is an increasing trend in the data, then M1(n,t) > M2(n,t) and M1(n,t) - M2(n,t) will be positive. Consequently, b(t) will be positive. Conversely, if there is a decreasing trend in the data then M1(n,t) < M2(n,t) and M1(n,t) - M2(n,t) will be negative. Consequently, b(t) will be negative. The relationship between M1(n,t) and M2(n,t), determined by the direction and magnitude of trend present in the data, in turn determines the sign of the slope of the double moving average forecast equation.

If we rewrite the equation for a(t) as a(t) = M1(n,t) + M1(n,t) - M2(n,t) we can better see the relationships between a(t) and the moving averages. If the trend is increasing M1(n,t) will fall behind the trend line in value. But M1(n,t) - M2(n,t) corrects for the lag. Conversely, if the trend line is decreasing, M1(n,t) will exceed the trend line in value. But M1(n,t) - M2(n,t) will be negative, and adding this negative value to M1(n,t) corrects for the lag.

If we were using our forecast equation FM2(n,t,t+i) = a(t) + b(t) *i to forecast the present, i would equal zero and FM2(n,t,t) = a(t). a(t) is a double moving average forecast of the present from the present. As we move through time, a(t) would have to be adjusted to account for the trend in our data. The amount of the adjustment is determined by the relationship between m1(n,t) and M2(n,t). a(t) represents the y-intercept of a straight line equation as the y-axis is being shifted along the time axis to represent the passage of time. Figure 10-6 depicts the phenomenon of our moving a(t).

Figure 10-6

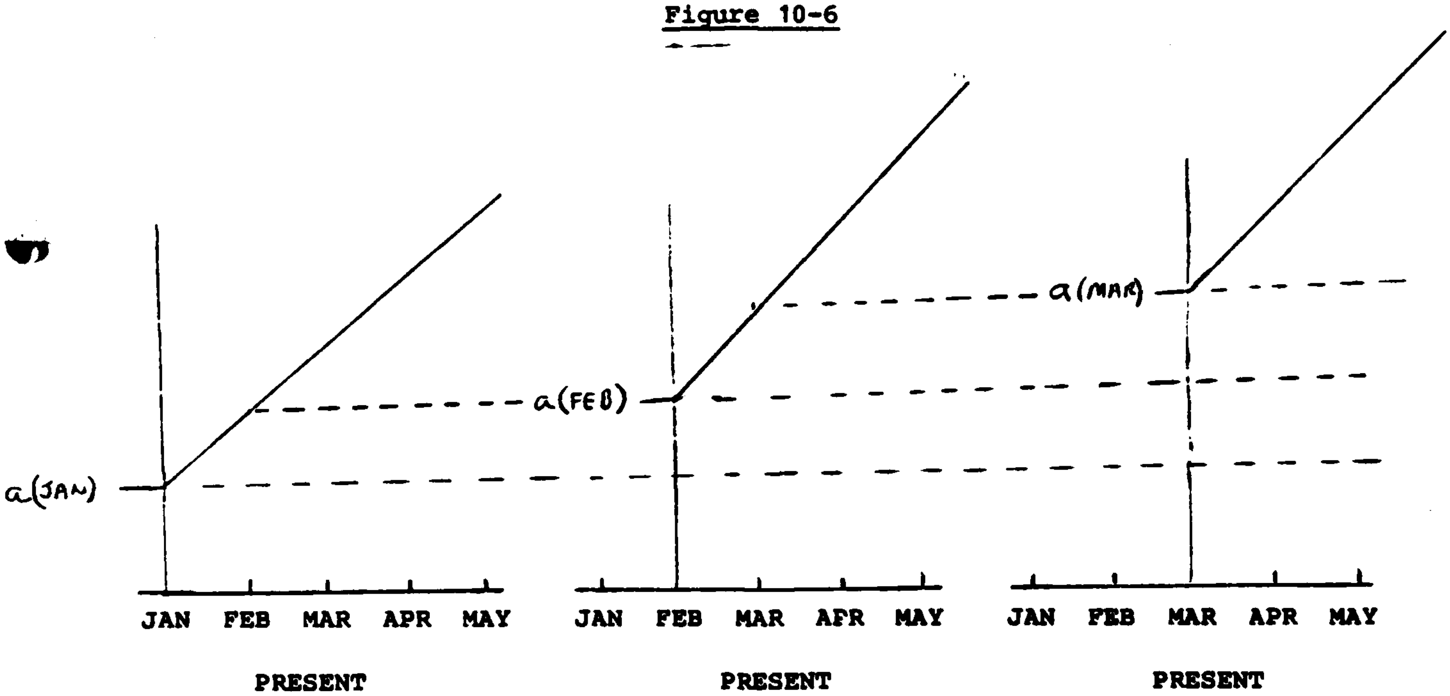


Figure 10-6 also implies that as the present moves forward in time by one period, the change in a(t) should equal the slope of the line. In other words, a(Mar) = a(Feb) + b and a(Apr) = a(Mar) + b. This phenomenon occurs because we are always forecasting from the present; the present always corresponds to i=o in our double moving average forecast equation.

When we applied our single moving average forecast to data with a linear trend our forecast was consistently biased. A single moving average forecast assumes a trend line with a slope of zero. Likewise, our double moving average forecast assumes a linear trend line. If our trend line is nonlinear, our double moving average forecast will also be biased. Furthermore, our double moving average forecast will move further and further away from the nonlinear trend line as the forecast interval increases. We do not have a forecasting technique employing moving averages that can be used to forecast nonlinear trends. However, you can use a double moving average forecast for a nonlinear trend provided that: (a) your forecast interval is relatively short, and (b) the curvature of your nonlinear trend line is relatively flat. Under these conditions a straight line can be used as a reasonable approximation of a nonlinear trend.

Throughout this section we have been assigning values for n on a rather arbitrary basis. We had stated earlier that the number of observations to include in the calculation of the moving average was a decision parameter for the forecaster. Now we need to ask, what is the criteria for selecting n?

We defined flexibility as the capability of a forecasting technique to respond to changes in the primary cause and effect relationships, i.e., changes in trend. For a forecasting technique to be flexible, it would have to be responsive to the most recent observations.

Moving average forecasts are flexible because the moving average forecast is continuously discarding older data and emphasizing new observations. The faster old data is discarded, the greater the influence new observations will have on the calculation of the forecasts, and the more flexible our moving average forecast will be. Therefore, the criteria of flexibility seems to dictate that old data should be discarded as rapidly as possible as new observations are entered into the data bank. In other words, n should be as small as possible. However, we also saw that a forecasting technique that is responsive to new observations will also be very sensitive to random fluctuations in the data. There is a tradeoff. If you want to be able to respond quickly to changes in trend, you have to be willing to accept relatively unstable forecasts. If you want stable forecasts relatively unaffected by random fluctuations in the data, you should select a large n.

The value you select for n will depend on three things: (1) your expectation that the future will bring a change in trend, (2) the cost or loss to you if you fail to forecast a change in trend, and (3) your ability to tolerate instability in your forecasts.

As you may have come to suspect, there is no one right answer to the problem of selecting n. The higher your expectation that the future will bring a change in trend, the greater your risk of loss if you fail to respond to a change in trend, and the greater your tolerance for forecasts subject to random fluctuations, the smaller the n you can select. However, if the trend line does not change, using a small n may result in larger forecasting errors because of the influence of random fluctuations in the

data. Your n must not be so small that every time you add a new observation your forecast is subject to wide oscillations between extreme values. Alternatively, your n must not be so large that a series of new observations identifying a new trend would have little impact on your forecast. A rule of thumb is than n should be no smaller than $\frac{9+i}{3}$ and no larger than the smallest n required to meet your minimum requirement for stability in your forecasts.

Two essential characteristics of a moving average forecast are: (1) the forecast is based on a limited number of observations, and (2) every observation has an equal influence upon the forecast because every observation is given equal weight. Note, it is the characteristic of equal weighting of the observations that prevents us from increasing the number of observations without a corresponding loss of responsiveness in our forecasts. What we need is a systematic scheme for assigning unequal weights to our observations in a manner that allows us to use more data as the basis for our forecast without losing responsiveness.

One possible mathematical scheme for assigning weights is called exponential smoothing. Exponential smoothing is a technique for assigning exponentially declining weights to the observations to produce an exponential average. Therefore, an exponential average is a weighted average that employs exponential weighting. We will discuss single and double exponential averages and their use in forecasting. Along with these new techniques we need to introduce some new symbology.

c is the exponential smoothing constant.

E1(c,t) is a single exponential average calculated for time period t with smoothing constant c.

E2(c,t) is a double exponential average calculated for time period t with smoothing constant c.

FE1(c,t,t+i) is a single exponential forecast made in time period t for time period t+i using smoothing constant c.

FE2(c,t,t+i) is a double exponential forecast made in time period t for time period t+i using smoothing constant c.

We stated in the introduction to this chapter than an exponential average is a weighted average that employs exponential weights. A weighted average of two observations would be $\overline{WY}$=W(1)*Y(1)+W(2)*Y(2), where W(1)+W(2)=1. The equations for all exponential averages are equivalent to this two-observation weighted average.

Let us look first at the equation for a single exponential average. Two substitutions are required. First, substitute the observed value of your variable for the time period in which you are calculating the exponential average in place of Y(1). Let Y(1)=Y(t). Next, substitute the single exponential average for the previous time period in place of Y(2). Let

Y(2)=E1(c,t-1). What about the weights? Let W(1)=c, our exponential smoothing constant. Since your weights should sum to one, let W(2)=1-W(1)=1-c. Our weighted moving average equation now becomes the equation for a single exponential average.

$$E1(c,t)=c*Y(t)+(1-c)*E1(c,t-1) \qquad \text{(Equation 10-1}_\text{J}\text{)}$$

A single exponential average is a weighted average of our most recent observation and the previous single exponential average.

But where is the exponential weighting? There are no exponential weights in Equation 10-13. To observe the exponential weighting present in Equation 10-13 we must first expand the equation by several terms. Our original equation is:

$$E1(c,t)=c+Y(t)+(1-c)*E1(c,t-1)$$

But Equation 10-13 tells us that:

$$E1(c,t-1)=c*Y(t-1)+(1-c)*E1(c,t-2)$$

Substituting into the first equation we get

$$E1(c,t)=c*Y(t)+(1-c)*\left[c*Y(t-1)+(1-c)*E1(c,t-2)\right]$$

or

$$E1(c,t)=c*Y(t)+c*(1-c)*Y(t-1)+(1-c)^2*E1(c,t-2)$$

But Equation 10-13 tells us that:

$$E1(c,t-2)=c*Y(t-2)+(1-c)+E1(c,t-3)$$

Substituting again we get

$$E1(c,t)=c*Y(t)+c*(1-c)*Y(t-1)+(1-c)^2*\left[c*Y(t-2) + E1(c,t-3)\right] \text{ or}$$
$$E1(c,t)=c*Y(t)+c*(1-c)*Y(t-1)+c*(1-c)^2*Y(t-2)+(1-c)^3*E1(c,t-3)$$

Since $(1-c)^0 = 1$, we can imagine that the coefficient for Y(t) is really $c*(1-c)^0$. Making this assumption, Table 10-23 presents each of the observations in the above equation and its corresponding coefficient.

Table 10-23

Observation	Coefficient
Y(t)	$c*(1-c)^0$
Y(t-1)	$c*(1-c)^1$
Y(t-2)	$c*(1-c)^2$

Can you guess what the coefficient for observation Y(t-3) would be? From the pattern which appears in our coefficients, your guess should be $c*(1-c)^3$. If you were to substitute Equation 10-13 for E1(c,t-3) into the above equation you would get

$$E1(c,t)=c*Y(t)+c*(1-c)*Y(t-1)+c*(1-c)^2*Y(t-2)+c*(1-c)^3*Y(t-3)+(1-c)^4*E1(c,t-4)$$

Your guess would have been correct. In fact, the coefficient for the general observation Y(t-k) in the above equation would be $c*(1-c)^k$. As you can see, the weight for each observation changes because of changes in the value of the exponent, k. Hence, exponential weighting. Each time we expanded our equation, the last term always turned out to be an exponential average. We could continue to substitute Equation 10-13 for the last term and expand our equation ad infinitum. In theory, the infinite sum that constitutes a single exponential average is

$$E1(c,t)=\sum_{k=o}^{\infty} c*(1-c)^k*Y(t-k) \qquad \text{(Equation 10-14)}$$

However, as a practical matter, an infinite series of observations is an impossibility. Therefore, we must modify our equation to fit the situation where we have only n observations.

$$E1(c,t)=\sum_{k=o}^{n-1} c*(1-c)^k*Y(t-k)+(1-c)^n*E1(c,t-n) \qquad \text{(Equation 10-15)}$$

To calculate the last term in Equation 10-15, E1(c,t-n), we need to know both Y(t-n) and E1(c,t-n-1). Since our data set runs from Y(t) to Y(t-n+1), we do not know the value of either of the above two terms. Therefore, we cannot calculate E1(c,t-n). We need to have some value for E1(c,t-n), however, in order to start the exponential smoothing process. Where will our value come from?

The solution to our problem is to assume a value for E1(c,t-n), and we have two alternatives upon which to base our assumption. Alternative one is to let our exponential average for time period t-n equal our observation for time period t-n+1, or E1(c,t-n)=Y(t-n+1). Equation 10-13 tells us that

$$E1(c,t-n+1)=c*Y(t-n+1)+(1-c)*E1(c,t-n)$$

If we let E1(t,t-n)=Y(t-n+1), then E1(c,t-n+1' will also equal Y(t-n+1). This will become the starting point for our exponential smoothing process. Alternative two would be to set our initial exponential average equal to $\bar{Y}(t)$. To illustrate both alternatives, let us look at the following data set.

Table 10-24

t	Y(t)
1	103
2	119
3	117
4	101
5	110

Let us arbitrarily set c=.20. Under our first assumption E1(.2,.0)=Y(1)=103. Then E1(.2,1)=(.2)*(103)+(8)*(103)=103. Which is what we knew would happen and,

E1(.2,2)=(.2)*(119)+(.8)*(103.00)=106.20
E1(.2,3)=(.2)*(117)+(.8)*(106.20)=108.36
E1(.2,4)=(.2)*(101)+(.8)*(108.36)=106.89
E1(.2,5)=(.2)*(110)+(.8)*(106.89)=107.51

Under our second assumption, $\bar{Y}(t) = \frac{550}{2} = 110 = E1(c,t-n)$.

Therefore:
E1(.2,1)=(.2)*(103)+(.8)*(110)=108.60
E1(.2,2)=(.2)*(119)+(.8)*(108.60)=110.68
E1(.2,3)=(.2)*(117)+(.8)*(110.68)=111.94
E1(.2,4)=(.2)*(101)+(.8)*(111.94)=109.75
E1(.2,5)=(.2)*(110)+(.8)*(109.75)=109.80

Table 10-25 contains the single exponential averages calculated under the two assumptions and the difference between the two averages.

Table 10-25

t	Y(t)	E1(.2,0)=103 E1(.2,t)	E1(.2,0)=110 E1(.2,t)	Difference
1	103	103.00	108.60	5.60
2	119	106.20	110.68	4.48
3	117	108.36	111.94	3.58
4	101	106.89	109.75	2.86
5	110	109.80	107.51	2.29

A look at the differences between the two single exponential averages should reveal two key points. First, the alternative you select for assuming a value for E1(.2,0) makes a difference in the calculated values of your exponential averages. Second, the difference between the two exponential averages decreases as you move further and further away from your starting point. Let us look again at Equation 10-15.

$$E1(c,t) = \sum_{k=o}^{n-1} c*(1-c)^{k}*Y(t-k)+(1-c)^{n}*E1(c,t-n)$$

The only difference between the two exponential averages in Table 10-25 is in the last term of the equation. When we assumed that E1(.2,0)=110, the last term of our equation became $(1-.2)^n*(110)=(.8)^n*(110)$. When we assumed that E1(.2,0)=103, the last term of our equation became $(1-.2)^n *(102)=(.8)^n*103$. Therefore, the difference between the two exponential averages should be $(.8^n*110)-(.8^n*103)=.8^n*(110-103)=.8^n*7$.

Table 10-26

t	$(1-c)^n = (.8)^n$	$.8^n*7$
1	$(.8)^1=.800$	5.60
2	$(.8)^2=.640$	4.48
3	$(.8)^3=.512$	3.58
4	$(.8)^4=.410$	2.87
5	$(.8)^5=.328$	2.30

A comparison between $.8^n*7$ and the Difference column in Table 10-25 shows the numbers to be almost equal. In fact, the numbers should be exactly equal, but there is some error in the calculations caused by rounding. The important point is that the difference between the two exponential averages is declining at an exponential rate, i.e., $(1-c)^n$. By the time you get to time period t=10, the difference between these two exponential averages will be $.8^{10}*7=0.75$.

The lesson to be learned is that if you are going to calculate single exponential averages you should have an adequate number of data points to continue your calculations out far enough where the effect of your assumed value for E1(c.t-n) on the value of your exponential average is negligible. Since E1(c,t-n) is weighted by the coefficient $(1-c)^n$, the smaller your smoothing constant the greater the weight assigned to E1(c,t-n) and the more observations you need to reduce $(1-c)^n$ to a small number. Table 10-27 gives you the smallest number of observations required to reduce the weight assigned to E1(c,t-n) to a value less than β for different values of β.

Table 10-27

Smallest values for n such that $(1-c)^n \leq \beta$

c \ β	.20	.15	.10	.05	.01	.001
.1	16	18	22	29	44	66
.2	8	9	11	11	21	31
.3	5	6	7	7	13	20
.4	4	4	5	5	9	14
.5	3	3	4	4	7	10

As a general rule, you would want the weight assigned to your assumed value for E1(c,t-n) to be .01 or less. Anytime the weight assigned to E1(c,t-n) becomes greater than .10 you have a potential source of distortion that could introduce significant error in your forecast. If you do not have a sufficient number of observations you can still calculate exponential averages. However, you should recognize in your analysis that your results may be distorted by the combined effects of your assumed value for E1(c,t-n) and your limited data.

Our single exponential average is very much like a single moving average. The primary difference is in the weighting scheme. Consequently, we would expect to find that the relationship between a single exponential forecast and a single exponential average parallels the relationship between a single moving average forecast and a single moving average. Equation 10-7 describes the latter relationship.

FM1(n,t,t+i) = M1(t,t+i) (Equation 10-7)

Our single exponential forecast is similarly determined by our single exponential average.

FEI(c,t,t+i)=E1(c,t) (Equation 10-16)

In words, our single exponential forecast made in time period t for time period t+i is the single exponential average made in time period t.

For an example, let us reexamine the data set used earlier to calculate single moving averages. The data, presented in Table 10-14 is reproduced in Table 10-28.

Table 10-28

t	Y(t)
1	74
2	94
3	84
4	71
5	88
6	81
7	77
8	91
9	87
10	80

Earlier we analyzed the above data for trend and found that at the 95% level of confidence we could accept the hypothesis that RS=0, that there is no trend. Therefore, let us begin our analysis by assuming that $E1(c,o)=\overline{Y}(t)=\frac{827}{10}=82.7$

For our smoothing constant let us set c = .25.

Then E1(.25,1)=(.25)*(74)+(.75)*(82.7)=80.525
and E1(.25,2)=(.25)*(94)+(.75)*(80.525)=83.894
and E1(.25,3)=(.25)*(84)+(.75)*(83.894)=83.921
and E1(.25,4)=(.25)*(71)+(.75)*(83.921)=80.691

For practice, you calculate E1(.25,5) for t=5 through t=10. Your results should look like Table 10-29.

Table 10-29

t	Y(t)	E1(.25,t)	FE1(.25,t,t+2)	EFE1:.25,t,t+2)
1	74	80.525	-	-
2	94	83.894	-	-
3	84	83.921	80.525	3.475
4	71	80.671	83.894	12.894
5	88	82.518	83.921	4.079
6	81	82.139	80.691	0.039
7	77	80.854	82.518	-5.518
8	91	83.391	82.139	8.861
9	87	84.293	80.854	6.146
10	80	83.220	83.391	-3.391

We measured the accuracy of our single moving average forecast by looking at the last six observations and calculating the MAD and the standard error of the forecast. Since our early single exponential averages are somewhat distorted by our assumed value for E1(.25,0), and to make our measures of forecasting error compatible, let us also calculate the MAD and the standard error of the forecast for FE1(.25,t,t+2) based on observations t=5 through t=10.

	Forecast	
Statistics	FMI(3,5,5+2)	FE1(25,t,t+2)
MAD	4.833	4.717
Standard Error	5.6421	5.400

Our single exponential forecast appears to be slightly more accurate than our single moving average forecast. However, the differences are so small as to be almost insignificant.

Our discussion of single exponential averages began by describing a single exponential average as similar to a single moving average using exponential weights. As a consequence of their similarities, you might suspect that the two averages would exhibit similar characteristics. One characteristic of a single moving average is that it produces biased forecasts when the data contains an increasing or decreasing trend. This same bias is also present in forecasts based upon single exponential averages. If there is an upward trend in the data, your forecasts will be consistently too low and if there is a downward trend in the data, your forecasts will be consistently too high.

You were able to solve the problem of forecasting with moving averages when there is a linear trend in the data by incorporating double moving averages into the calculations of your forecast. Your forecast was based upon the parameters a(M2) and b(M2), which in turn were dependent upon the values M1(n,t) and M2(n,t). However, the equations for a(m2) and b(M2) are dependent upon the fact that the observations are equally weighted.

Because exponential weighting represents a systematic scheme of weighting with known and consistent mathematical properties, equations have also been developed that will allow us to use double exponential averages as a basis for a forecast when there is a linear trend in the data.

Before you can calculate your forecast, however, you must first calculate the double exponential average. In the discussion of single exponential averages, the statement was made that the equations for all exponential

averages are equivalent to the two-observation weighted average equation MW=W(1)*Y(1)+W(2)*Y(2). Bv making several substitutions in the above equation, Equation 10-13 for a single exponential average was derived.

$$E1(c,t)=c*Y(t)+(1-c)*E1(c,t-1) \qquad \text{(Equation 10-13)}$$

A double exponential average is also a weighted average. The weights are still c and 1-c. However, a double exponential average is a weighted average of the single exponential average in time period t and the double exponential average in the previous time period.

$$E2(c,t)=c*E1(c,t)+(1-c)*E2(c,t-1) \qquad \text{(Equation 10-17)}$$

If you were to go through the same exercise expanding Equation 10-17 that you did with Equation 10-13, you would find that Equations 10-14 and 10-15 have parallel counterparts for double exponential averages.

For a single exponential average, Equation 10-14 says

$$E1(c,t) \sum_{k=o}^{\infty} c*(1-c)^k*Y(t-k) \qquad \text{(Equation 10-14)}$$

The parallel equation for a double exponential average is

$$E2(c,t) \sum_{k=o}^{\infty} c*(1-c)^k*E1(c,t-k) \qquad \text{(Equation 10-18)}$$

Therefore, a double exponential average is an exponentially weighted sum of single exponential averages in the same way that a single moving average is an exponentially weight sum of the observations. Because you never have an infinite data set, you were forced to modify Equation 10-14 for a finite data set of n observations to read

$$E1(c,t)=\sum_{k=o}^{n-1} c*(1-c)^k*Y(t-k)+(1-c)^n*E1(c,t-n) \qquad \text{(Equation 10-15)}$$

Likewise, to account for the realism of finite data sets, Equation 10-18 must be modified to read

$$E2(c,t)=\sum_{k=o}^{n-1} c*(1-c)^k*E1(c,t-k)+(1-c)^n*E2(c,t-n) \qquad \text{(Equation 10-19)}$$

At this point you have the same problem you had with single exponential averages. Your data set consists of observations Y(1) through Y(n). However, to calculate E2(c,t-n) you need to know E1(c,t-n) and E2(c,t-n-1), and you know neither one. What are you going to do? When discussing single exponential averages you were forced to assume a value for E1(c,t-n), and, since there was no trend in the data, you chose to let E1(c,t-n)=Y(t). Now you must also assume a value for E2(c,t-n) if you are going to forecast with double moving averages. What value should we assume? To answer this question, you must first discuss the double exponential forecast.

The double exponential forecast, like the double moving average forecast, is used to forecast a time series variable when there is a linear

trend. Therefore, our forecast equation should look like the equation for a straight line, and it does.

$$FE2(c,t,t+i)=a(E2)+b(E2)*i \qquad \text{(Equation 10-20)}$$

Like the double moving average forecast, the parameters for our double exponential forecast a(E2) and b(E2). In fact, the equations are very similar. For our double exponential forecast, the equations are:

$$a(E2)=2*E1(c,t)-E2(c,t) \qquad \text{(Equation 10-21)}$$

$$b(E2)= \frac{c}{1-c} *\left[E1(c,t)-E2(c,t)\right] \qquad \text{(Equation 10-22)}$$

We finished our discussion of double exponential averages with the question of what values to assume for E1(c,t-n) and E2(c,t-n) left unanswered. Now that we have the equation for a(E2) and b(E2) upon which we will base our double exponential forecast, we can use these equations as a basis for developing our initial assumed values for our exponential averages.

Ordinarily, you will calculate E1(c,t) and E2(c,t) first, and then calculate a(E2) and b(E2). However, to begin our exponential smoothing process we are going to reverse the two steps. We are going to calculate estimates of a(E2) and b(E2) for time period t-n and then use these estimates to find our assumed values for E1(c,t-n) and E2(c,t-n). Where are we going to get our estimates for a(E2) and b(E2) in time period t-n? Answer, from a LSBF line.

An example would best describe the process. Suppose we were analyzing the data set in Table 10-30.

Table 10-30

t	Y(t)
1	27
2	30
3	31
4	32
5	34
6	37
7	40
8	41
9	43

If you were to fit a LSBF line through this data you would come up with the formula Y(t)=25+2*t. From our LSBF line you extract the two parameters, a=25 and b=2. These two parameters become your estimate for a(E2) and b(E2) for time period t-n. Substituting into Equations 10-21 and 10-22.

$$25=2*E1(c,t-n)-E2(c,t-n)$$

$$2= \frac{c}{1-c} *\left[E1(c,t-n)-E2(c,t-n)\right]$$

All that remains is to select our smoothing constant and solve for E1(c,t-n) and E2(c,t-n). For the sake of this example, let us set c=.2. Based on our LSBF parameters a=25 and b=2 we can calculate our assumed values for E2(c,t-n) and E2(c,t-n) which are 17 and 9 respectively. We can now begin to calculate our exponential averages.

Table 10-31

t	Y(t)	E1(.2,t)	E2(.2,t)
0	27	17*	9*
1	30	19.00	11.00
2	31	21.2	13.04
3	32	23.16	15.06
4	34	24.93	17.03
5	37	26.74	18.97
6	40	28.79	20.93
7	41	31.03	22.95
8	43	33.02	24.96
9		35.02	26.97

*Assumed values based on our LSBF line

Given a value for E1(.2,t) and E2(.2,5) *you* would calculate a(E2) and b(E2) for each time period. You would then base your linear forecast on these two parameters. In our example, suppose you were forecasting 3 periods into the future. Your forecast and your forecasting error would look like Table 10-32.

Table 10-32

t	Y(t)	a(E2)	b(E2)	FE2(.2,t,t+3)	EFE2:.2,t,t+3)
1	27	27.00	2.00	-	-
2	30	29.36	2.04	-	-
3	31	31.26	2.03	-	-
4	32	32.83	1.98	33.00	-1.00
5	34	34.51	1.94	35.48	-1.48
6	37	36.65	1.97	37.45	- .35
7	40	39.11	2.02	38.77	1.23
8	41	41.08	2.01	40.33	.67
9	43	43.07	2.01	43.56	.44

The values of b(E2) in Table 10-32 reflect a basic characteristic of both double moving average forecasts and double exponential forecasts. The value of b(E2) will fluctuate around the true slope of the trend line. However, if the slope of the trend line changes, b(E2) possesses the ability to adjust to the change.

To impress this point upon you, let us look at a set of data that has a change in trend. Table 10-33 consists of nine observations. The first observation is from the trend line Y(t)=100+5*t. Observation three through nine are from the trend line Y(t)=70+20*t. Observation two is the intersection of the two trend lines. For our initial estimates set a(E2)=a=100 and b(E2)=b=5 from the first trend line. From these estimates assumed values of E1(c,o) and E2(c,o) can be calculated. For a smoothing constant c=2/7, E1(c,o) and E2(c,0) equal 87.5 and 75 respectively. Table 10-33 demonstrates how the slope of a double exponential average adjusts for the change in trend.

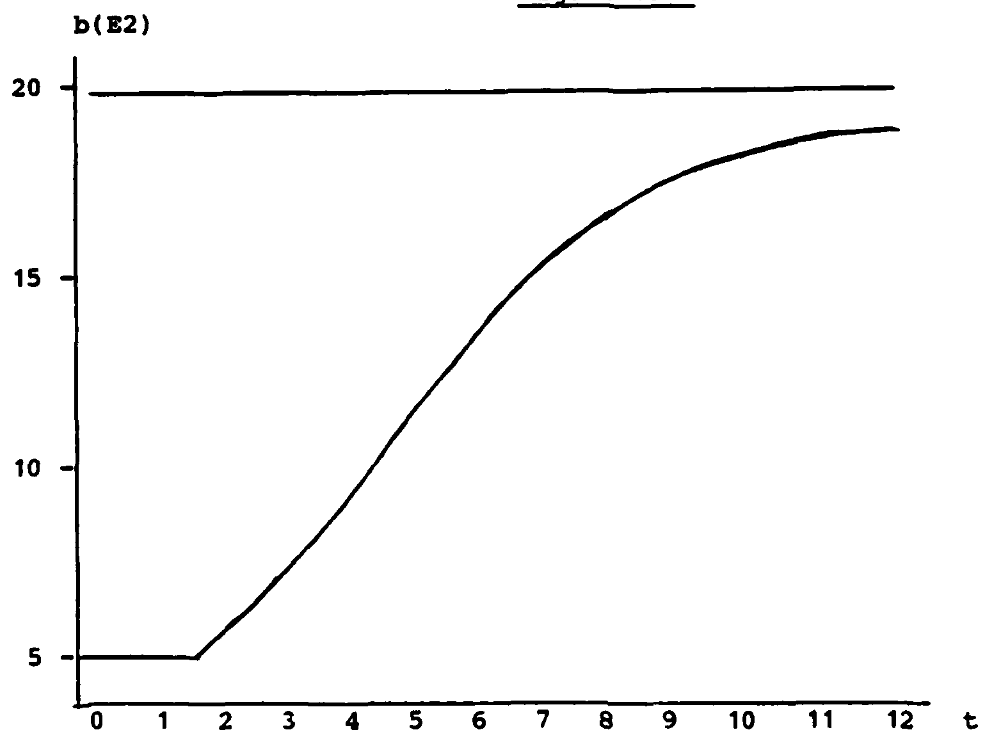
Table 10-33

t	Y(t)	E1(2/7,t)	E2(2/7,t)	b(E2)	Δb(E2)
0	-	87.500	75.000	5	-
1	105	92.500	80.000	5	0
2	110	97.500	85.000	5	0
3	130	106.786	91.225	6.224	1.224
4	150	119.113	99.199	7.974	1.750
5	170	133.666	109.047	9.848	1.874
6	190	149.761	120.680	11.632	1.784
7	210	166.972	133.906	13.226	1.594
8	230	184.980	148.499	14.529	1.366
9	250	203.557	164.230	15.731	1.139
10	270	222.541	180.890	16.660	.929
11	290	241.815	198.297	17.407	.747
12	310	261.296	216.297	18.000	.593

Several characteristics of b(E2) can be detected by examination of Table 10-33. First, it is evident that b(E2) is responding to the change in the trend line. In the ten time periods from time period 2 to time period 12, b(E2) has more than tripled from 5 to 18. Second, b(E2) responded immediately to the change in trend. Time period 3 is the first time period not on the old trend line, and b(E2) increased from 5 to 6.224. Take note, however, that b(E2) would have increased even if Y(3) had been solely a random fluctuation and not the beginning of a new trend. Third, b(E2) increased at an increasing rate for the first three observations after the change in trend. In the ten time periods from time period 2 to time period 12, b(E2) increased by 260%. However, half that increase was realized by time period 6, 4 time periods after the change in trend. Last, after time period 5, b(E2) begins to increase at a decreasing rate. In fact, this trend will continue as b(E2) approaches the slope of the new trend line asymptotically. b(E2) will never equal 20 because the observations from the old trend line will continue to affect the calculations of b(E2) no matter how small their exponential weights become. If you were to graph

b(E2), your graph would look as follows:

Figure 10-7



Our forecast for any future time period depends upon the value of a(E2) and b(E2). a(E2) and b(E2), in turn, change as we move from time period to time period. Therefore, we would expect that our forecast of a future time period will also change as we move through time. For example, let us compare our forecast of time period t=9 as we move from time period t=1 to t=8. To do this let us use our data from Table 10-34.

Table 10-34

t	a(E2)	b(E2)	FE2(2/7,t,9)	E(FE2:2/7,t,9)
1	27.00	2.00	43.00	0
2	29.36	2.04	43.64	-.64
3	31.26	2.03	43.44	-.44
4	32.83	1.98	42.73	.27
5	34.51	1.94	42.37	.73
6	36.65	1.97	42.56	.44
7	39.11	2.02	43.15	-.15
8	41.08	2.01	43.09	-.09

Since we know that Y(9)=43, we can calculate the error in our forecast as we move through time. Note that even though our error was at a minimum in time period 8, our forecast did not necessarily improve as we moved forward in time. Our forecast was more accurate in time period 2 than it was in time period 5, and more accurate in time period 4 than either time period

5 or time period 6. It is important to recognize that while your most recent forecast may be your "best" forecast in terms of the available data, it may not be your most accurate forecast.

We began our discussion of single exponential averages by showing that E1(t,c) was nothing more than a weighted average.

$$E1(c,t)=c*Y(t)+(1-c)*E1(c,t-1) \qquad \text{(Equation 10-13)}$$

The sum of the weights are equal to c+(1-c)=1. Then we expanded our equation for E1(t,c) into an infinite exponentially weighted sum of past observations.

$$E1(t,c)=\sum_{k=o}^{\infty} c*(1-c)^{k}*Y(t-k) \qquad \text{(Equation 10-14)}$$

Does the sum of our exponential weights also add to one? The answer is yes, provided that our smoothing constant, c, is less than or equal to one and greater than or equal to zero.

$$\sum_{k=o}^{\infty} c*(1-c)^{k} =1 \text{ for } 0 \leq c \leq 1 \qquad \text{(Equation 10-23)}$$

Thus we can see that our single exponential average is equivalent to a weighted average where the weights are exponential, they sum to one, and the number of observations in the weighted moving average is infinite. We then modified our equation to account for finite data sets.

$$E1(t,c)=\sum_{k=o}^{n-1} c*(1-c)^{k}*Y(c-k)+(1-c)^{n}*E1(c,t-n) \qquad \text{(Equation 10-15)}$$

Do our weights still add up to one? The answer, mathematically speaking, is no.

$$\sum_{k=o}^{n} c*(1-c)^{k} = 1-(1-c)^{n+1}$$

However, you must remember that theoretically the term E1(c,t-n) is itself a summation of exponentially weighted observations. The weights c*(1-c) for k=n+1 to infinity are all incorporated in the term E1(c,t-n). And

$$\sum_{k=n+1}^{\infty} =c*(1-c)^{k}=(1-c)^{n+1}$$

Therefore, theoretically, all of the weights are present in Equation 10-15 and they still sum to one provided that our smoothing constant, c, is less than or equal to one and greater than or equal to zero.

The question now is, given the restriction that $0 \leq c \leq 1$, how do we select a particular value for our smoothing constant. The component of our exponential weight that declines as you move backward in time is $(1-c)^k$. This term declines slowly when c is very small, and quickly when c is very large. Table 10-35 demonstrates how $(1-c)^k$ behaves for three different values of c, .95, .5, and .05.

Table 10-35

k	(1-.95)	(1-.5)	(1-.05)
1	.0500	.500	.95
2	.0025	.250	.903
3	$.125*10^{-3}$	.125	.857
4	$.625*10^{-5}$	$.625*10^{-1}$	.815
5	$.313*10^{-6}$	$.313*10^{-1}$	.774
6	$.156*10^{-7}$	$.156*10^{-1}$	.735
7	$.800*10^{-9}$	$.781*10^{-2}$	.698
8	$.391*10^{-10}$	$.391*10^{-2}$	.663
9	$.195*10^{-11}$	$.195*10^{-2}$	.630

There is one important comparison that can be made examining Table 10-35. First, a small c discounts the past at a very slow rate. Older observations can still have a relatively significant effect on the exponential average, and consequently, on the forecast. Compared to moving averages, a small c corresponds in its effects to a large n for a moving average. Conversely, a large c discounts the past at a very rapid rate. Older observations have a negligible impact on the exponential average, and consequently on the forecast. A large c corresponds in its effects to a small n for a moving average.

Therefore, your considerations in Chapter 3 regarding the selection of n for a moving average forecast also apply to the problem of selecting c for an exponential average forecast. If you select a large c, your forecast will be very responsive to changes in trend, but also subject to distortion by random fluctuations in the data. If you select a small c your forecast will be relatively unaffected by random fluctuations in the data, but it will also be unresponsive to changes in trend. It is the same tradeoff that we have been confronting throughout this chapter.

The value you select for c will depend on the same three factors that influenced your selection of n for a moving average: (a) your expectation that the future will bring a change in trend, (b) the cost or loss to you if you fail to forecast a change in trend, and (c) your ability to tolerate instability in your forecasts. The greater your expectations that the future will bring a change in trend, the greater your risk of loss if you fail to respond to a change in trend, and the greater your tolerance for instability in your forecast, the larger the smoothing constant you can select.

A large c, however, makes your forecast sensitive to random fluctuations in the data. The further out into the future you are trying to forecast, the greater the magnitude of the potential distortion that can occur because of random fluctuations. Therefore, you must place limits on the maximum value you select for c based upon the number of time periods into the future you are trying to forecast. A general guide is to restrict your chosen value for c to the range 0 to $\frac{2}{i+1}$. Table 10-36 lists representative maximum values for smoothing constants based on this rule.

Table 10-36

i	$0 \leq c \leq \frac{2}{i+1}$
1	1.000
2	.670
3	.500
4	.400
5	.330
6	.286
7	.250
8	.222
9	.200
10	.182

If you select a value for your smoothing constant that is larger than this rule would indicate, be aware that your forecasts may be distorted by the random fluctuations present in your most recent observations. Furthermore, the further out into the future you forecase, the greater the effect this distortion will be.

CHAPTER 11

COST ESTIMATION

INTRODUCTION

Cost estimating or cost prediction encompasses a wide variety of tasks and techniques which, individually, may be quite simple, but taken as a whole, become quite complex. For example, estimating the costs for a total force structure is more involved than individual system costing because the former must take into account the interactions of all the individual systems which make up the total force.

In the acquisition process, costs are normally categorized as:

Research and Development costs: All costs associated with bringing a system into production configuration. These costs include concept studies, engineering design, program management, advanced development, and may even include costs for construction of prototypes.

Investment costs: All costs associated with phasing a new system into the operational inventory. These include manufacturing or production, construction of facilities, initial spares and support equipment, technical data, initial training, and other initial logistic support costs.

Operating and Support costs: All costs associated with operating and maintaining a system after it has been phased into the operational inventory. Whereas Research, Development and Investment costs are "one-time" expenditures, Operating and Support costs are recurring throughout the system's life cycle.

Costs may also be viewed as incremental when they account for the amount of additional resources to be consumed after the new system is in place, including consideration for any old system phase-outs, manpower realignment, and resource reallocation. In essence, incremental costing ignores sunk costs and is concerned only with changes in cost of a new system over and above costs already being incurred on the old system.

AFSCM 173-1, Cost Estimating Procedures, describes the cost estimating process as one which involves five steps.

1. Define and plan the task.
2. Select the estimating structure.
3. Collect, evaluate, and adjust the necessary cost and cost related data.
4. Select and apply the estimating method.
5. Document the analysis and results.

Although each of the steps warrants special consideration, the emphasis here will be directed toward steps 3 and 4. It is important to note that even though data collection and selection of estimating methodology are listed as separate tasks, to do one without full consideration of the other would result at best in a haphazard effort. For example, it would be just as inappropriate to formulate a cost model without consideration of available (or useable) data as it would be to manufacture an airframe without considering the engine(s) to propel it.

COSTING METHODS

Except when used to develop budget estimates which later become ceilings on expenditures, most costing methodologies are directed toward reducing uncertainty, i.e., to determine affordability or to aid in the decision between competing alternatives. Several of the most commonly used methods are given below.

Industrial Engineering. The industrial engineering method for estimating the cost of a system or component consists of aggregating part by part estimates to get total system cost including the cost of assembly. This method requires that the actual system final configuration be known. Obviously, the industrial engineering approach cannot be used for cost estimating in the conceptual and early development phases where engineering designs are in a state of flux. Furthermore, because of the large number of separate estimates to be aggregated, this approach is often cumbersome and costly to perform. And finally, even when system (or component) configuration is known, it is often very difficult to obtain the necessary cost data.

Analogy. Analogy cost estimating can be used for costing equipment items when technical specifications are available and the specifications and costs are known for a comparable item. Ordinarily, analogy cost estimating is used when:

1. A quick study is needed;
2. Estimates are needed for low-cost system components; or,
3. Where it is not practical to secure a large sample of past costs and technical data.

To apply this method the analyst must first identify a suitable "old" counterpart for the new item. Then an appropriate technical characteristic must be chosen, since this approach assumes that cost is proportional to the magnitude of some technical characteristic. The following example illustrates how the weight of an item may be used to estimate the cost of a new component.

$$C_n = C_o \times \frac{W_n}{W_o}$$

where C_n = estimated cost of new component
C_o = actual cost of a comparable component
W_n = weight of the new component
W_o = weight of the old component

Although its usefulness should not be ignored, analogy cost estimating has a serious limitation. It uses only one observation as a basis for the estimate as well as only one explanatory variable, which is weight in the previous example. If unusual circumstances are associated with either the new or old component, the resulting estimate may be severely biased.

The major advantage of analogy costing is that it is quick and easy to apply. In addition, it is often used to obtain "rough" initial estimates or as a check of the reasonableness of other more complex methods of costing.

Cost Estimating Relationships. One of the most useful cost estimating techniques is the cost estimating relationship (CER). The CER is developed from historical data from a similar system or component to predict costs for the new system. CERs normally relate cost as the dependent variable to performance parameters or other system characteristics, the independent variables, in the form of a multiple regression model. See Chapter 13 for a discussion of regression analysis. This technqiue of relating system parameters to cost is commonly referred to as parametric costing.

Separate CERs are usually developed for each of the cost categories which comprise the major system. The CERs are then summed to obtain the total system cost estimate. For example, the following CER was developed to estimate the avionics cost per unit for a fighter aircraft:

$$C = 3.296 + .05185X_1 + .0305X_2 - .00555X_3 + .0213X_4 - .003704X_5$$

where C = Production cost of the 150th unit in millions of 1975 dollars

X_1 thru X_5 = Complexity factors derived through other CERs involving characteristics of the aircraft performance and design, such as gross take-off weight, cruising speed, number of crew members, thrust, etc.

Although the example above is a linear function and lends itself very well to least squares regression techniques, many CERs can and have been developed using nonlinear curve fitting techniques as in the following example estimating avionics production costs for a cargo aircraft.

$$C = (1.7926 \times 10^{-9}) X_1^{2.540} \bullet X_2^{1.340} \bullet X_3^{.960}$$

where C = Production cost of the 150th unit in millions of 1975 dollars.

X_1 = First flight year

X_2 = Total quantity of avionics subsystems per aircraft

X_3 = Length of aircraft fuselage

However, an equation of this form can be transformed into a linear equation by:

$$\log C = \log(1.7926 \times 10^{-9}) + 2.54 \log X_1 + 1.34 \log X_2 + .96 \log X_3$$

The manager should pay particular attention to the fact that any CER provides a single, point estimate of cost. This is an <u>average</u> cost, sometimes referred to as a "single best guess". It is usually much more beneficial to provide the user with a range, or bounded estimate of costs in the form of a prediction interval. A prediction interval enables the user to make statistical statements about the costs estimated by the CER. For example, one might say that: "I am 95 percent confident that the cost of the new engine will be between \$409,000 and \$621,000", as opposed to using the point estimate of \$515,000. This becomes extremely important, if not critical, when the CER itself indicates a weak relationship among the dependent and independent variables. After all CERs and corresponding prediction intervals have been developed, the cost estimated by each are then summed to find the total system cost. Although the point estimates can be totalled directly, prediction intervals cannot. The Rand Corporation has developed a technique for constructing prediction intervals for these summed totals. The application of the technique is dependent upon the variance (amount of error) which exists in the individual CERs. A complete discussion of this technique can be found in the Rand Corporation memorandum RM-5806-PR, October 1968.

<u>Cost Analysis</u>. Useful applications of the techniques presented in this chapter are found in the various methods of cost analysis. Indeed, most forms of cost analysis involve cost estimates of some kind. The objective of cost analysis is to use cost as an input in the decision making process. In this regard, cost estimates add vital information and help to reduce the uncertainty surrounding the choice among alternatives. Cost analysis has assumed an increasingly more important role in the current era of tight defense budgets, rising costs, and scarce resources. In the Department of the Air Force, the responsibility for cost analysis falls under the Directorate of Cost and Management Analysis, Comptroller of the Air Force. Actual analyses, however, are accomplished at all levels of Air Force operations. For whatever purpose, to support a budget estimate, a base closing, or a contract for janitorial services, cost analysis provides a valuable input in the decision making process. Depending upon its intended purpose, a cost analysis may take on one of the specific forms described below.

<u>Benefit-Cost Analysis</u>. One of the techniques of cost analysis which has been used by the Army Corps of Engineers is benefit-cost analysis. Its primary purpose is to show a relationship between the costs incurred in a specific project and the benefits to be derived from that project. This technique has frequently been used to analyze (and sometimes justify) Federal water resources projects such as dams. In a benefit-cost analysis, costs which will be incurred (C) are combined with the benefits to be achieved (B) over a specified period of time in the following ratio:

$$\frac{B}{C}$$

When the ratio exceeds 1, obviously, benefits exceed costs; which would indicate the project may be worthwhile. When two or more alternatives are being considered, the project with the highest benefit-cost ratio would prevail, in the absence of other decision criteria. In addition, the benefit-cost ratio could be used to rank projects which are not mutually exclusive. This would give the decision maker an indication of which projects would yield the greatest benefits in relation to their costs. This technique, however, is not without serious drawbacks. In the first place, it does not consider the magnitude of the costs/benefits involved. If the decision maker ranks all of the projects by B/C, the very first project may consume the entire budget, leaving nothing for other necessary programs. Secondly, significant difficulties arise when the analyst attempts to estimate downstream benefits, particularly long-term benefits. Despite its shortcomings, however, benefit-cost analysis has found some successful applications. Its use has been extended to such diverse areas as medical research and urban transportation problems.

Cost Effectiveness Analysis. This type of cost analysis is similar to the method discussed above, except that a measure of effectiveness is used as the benefits to be attained. For example, in comparing alternate weapon systems under consideration, the ratio of system availability versus life cycle cost could be formed as:

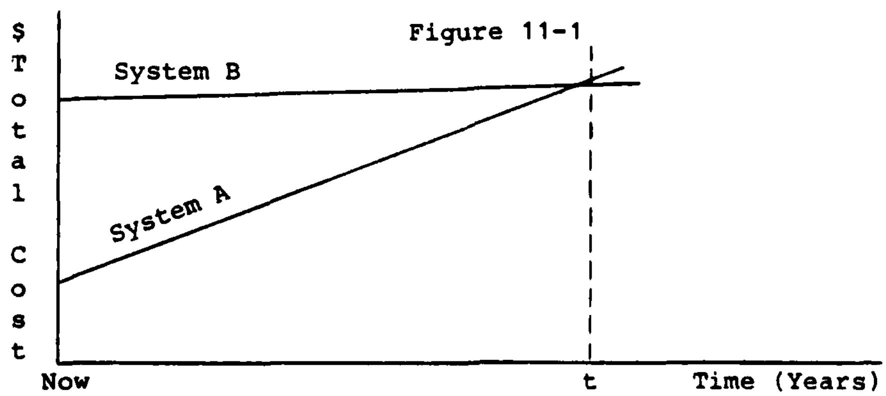
$$\frac{\text{Availability}}{\text{LCC}}$$

Other possibilities include: $\dfrac{\text{Mean-time-between-failure}}{\text{Life cycle costs}}$ or $\dfrac{\text{Reliability}}{\text{LCC}}$

In each case, the highest effectiveness/cost ratio might be basis which could be used by the decision maker to aid in the decision between alternatives. This technique is particularly useful when the life cycle costs of competing systems are roughly similar.

Break-even Analysis. This technique of cost analysis is especially useful in resolving the problem of equipment replacement. In break-even analysis, the analyst computes the cost stream of keeping the old system on hand and compares it with the cost profile of the proposed system as illustrated in the following chart.

Figure 11-1



Here, System A is the old system which will incur an increasing amount of operation and support costs as the system continues to age and wear out. The investment costs for System B, the proposed replacement, cause the total costs for System B to be higher initially. After year t, the break-even point, total system cost for A exceeds that of B. If the useful life of both systems is greater than t, then System B should be purchased now.

Although the example is over-simplified, the technique has been found to be valuable in replace versus modify decisions. Its utility, however, diminishes when high technology items with relatively short useful lives are considered.

Cost Sensitivity Analysis. One of the primary functions of cost estimating and cost analysis is to provide information to the user concerning the sensitivity to total costs to changes in assumptions or input parameters. For example, a sensitivity analysis could be performed to determine how a change in hours-between-removal of an engine affects the total maintenance costs for the engine. If the analysis shows that there is minimal effect, efforts to reduce engine maintenance costs can be directed elsewhere.

Since cost models are developed and used extensively in cost estimating, sensitivity analysis is performed by varying inputs to the model and examining the resulting cost estimate (output) for significant changes. Similarly, the assumptions which were made in model development should also be tested for cost sensitivity. For example, if an operating and support cost (O and S) model for an aircraft is based upon the assumption that annual flying hours will be 1,100 hours, then sensitivity analysis should be performed to determine the impact on O and S costs from changes in the annual flying hour program.

SPECIAL CONSIDERATIONS

Although the cost estimating technique used will depend largely upon the purpose for which the estimate is being developed, there are certain criteria which should be followed during any cost estimating process.

Data. Since the value of the cost estimate depends upon the accuracy of the cost data used, a great deal of care should be taken to insure that the data source(s) is/are reliable. Whenever possible, two sources should be used, one to verify the other.

Tracking. Cost estimates are often used to forecast or predict costs which will be incurred during some future interval of time. When this is the case, the estimator should compare the estimate with the actual costs when they occur. This procedure, called tracking, provides valuable information. If significant variances are detected between actual and estimated costs, the cause of the variance should be investigated. Was the estimate erroneous because of the estimating technique, data, or assumptions? The answers to the question can serve as valuable inputs to future cost estimates and will help to improve the process.

Present Value. Whenever a cost estimate is required as an input to the selection of a system or project, the costs (and savings) must be discounted at a rate of 10 percent, except:

1. Water resources projects,

2. Commercial services obtained under AFR 26-12, or

3. When the project or system has an expected useful life less than three years.

Methodology for applying the present value criterion is contained in Chapter 21.

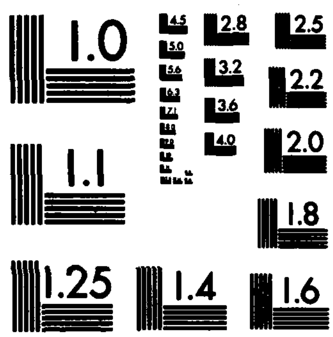
Inflation. When deriving cost estimates, costs should be presented and evaluated with constant and current (inflated) dollars. Specific indices for this purpose can be found in AFP 173-13, USAF Cost and Planning Factors, Vol I. When estimating future costs, dollars should first be inflated through the use of an approved index, and then discounted at the appropriate rate.

CONCLUSION

The techniques and methodologies described in this chapter merely scratch the surface on the use of costs in reducing uncertainties surrounding the decision process. The choice of the appropriate technique depends ultimately upon the objective of the analysis and the data available as inputs.

AD-A121 990 QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS J E ENGEL APR 80 AFIT-LS-32 4/7

UNCLASSIFIED F/G 15/5 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

CHAPTER 12

LIFE CYCLE COST

INTRODUCTION

The rapidly increasing costs of modern defense systems and even more rapidly increasing costs to operate and support them has been a major concern of Department of Defense (DOD) managers. Recognizing that downstream operating and support costs often are several times greater than the initial acquisition cost of a system, DOD managers have introduced the concept of life cycle cost into the systems development and acquisition process.

AFR 800-11 defines life cycle cost (LCC) as the total cost of an item or system over its full life, including development, acquisition, operation and support, and disposal. In essence, LCC represents the total cost of ownership of a system or component from "cradle to grave."

With the introduction of LCC, cost has been elevated as a major decision criteria, equal in importance with performance and schedule. Cost in this sense, of course, means life cycle cost. And in order to reduce total life cycle cost, the impact of various systems designs and configurations on operating and support costs must be carefully evaluated early in the system's design phases.

Figure 12-1 illustrates the relationships of the major categories of LCC to the phases of the systems life cycle. After Milestone O, the decision to proceed into the succeeding phase of system development must be supported by a thorough analysis of the LCC impact of alternate designs. The objective is to insure that development continues not only with a system which is needed, but one which can be afforded.

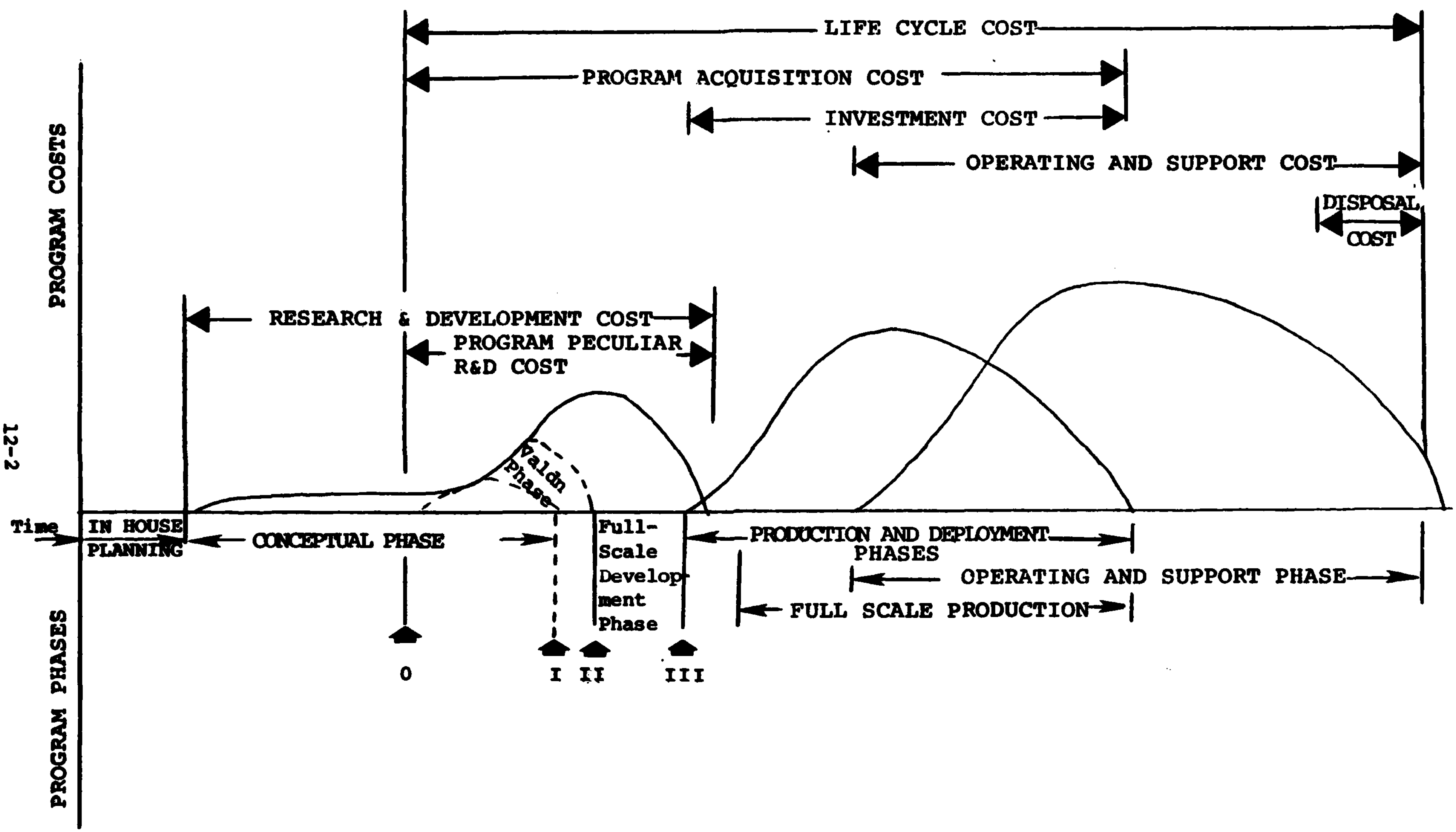
LCC APPROACHES TO SYSTEM ACQUISITION

Life cycle cost can be viewed as a three-sided approach to system cost management with the single purpose of acquiring systems which meet minimum operational requirements for the lowest overall cost.

In the first place, LCC is a procurement technique. Competing systems are evaluated on their total cost over their useful lives, rather than solely on their acquisition costs.

Secondly, LCC is an estimating technique. Comparisons are made among existing systems and those which have been proposed. In this respect, the costly lessons learned in the past play an important role in the selection and development of new systems. LCC also encourages early trade-off studies to determine the impact on total costs for increased performance, reliability, and maintainability.

Figure 12-1
SYSTEM LIFE CYCLE COST



And thirdly, LCC is used to develop design targets. Cost goals are established to control operating and support costs as well as acquisition costs. The term design-to-cost, formerly associated with acquisition costs only, has taken on added meaning so that now it is more accurately termed design-to-LCC. Design-to-LCC encourages design trades so that the overall cost impact of alternative system/component designs can be determined. In viewing total system cost, it could be beneficial to incur an increase in acquisition costs if the net result is lower downstream operating and support costs and ultimately, life cycle cost.

The general approach stipulated when developing new items/systems is to use the design-to-life-cycle-cost approach, commonly with a design to unit production cost goal and with contractual riders concerning ownership costs. This approach underlines the idea that cost has become a parameter equal in importance with performance and schedule. This requires techniques to estimate all LCC categories during the design phase, with sensitivity to design changes. This is not easy or straightforward and cannot be done with absolute accuracy, but techniques exist which are believed to give reasonable accuracy or "figures of merit" which can be used in the various decisions required, such as design trades, source selection, and structuring contractual incentives.

All standard cost estimating methods and techniques are used in life cycle costing. Considerable use is made of the analogy and parametric methods since they can be structured to use data available in the design phase. (See Chapter 11). The regression technique is widely used in estimating LCC (See Chapter 13).

LCC Categories

A typical life cycle cost breakdown structure is shown in Figure 12-2. All costs associated with the system(s) under consideration are classified under the three major cost categories: Research and Development, Investment, and Operations and Support. Such a categorization of costs can serve several purposes:

1. Data collection - cost data can be accumulated into specific "accounts" which have been identified as being the most useful.

2. Comparison of alternatives - when costs for competing systems are collected under standard formats, comparing these systems on a cost basis becomes an easier and more meaningful task.

3. Model building - with standard classifications of costs, models can be constructed from a readily available data base.

4. Trade-off analysis - a cost breakdown structure facilitates cost trade-offs among competing systems as well as identification of major cost drivers.

For a typical (but not all) weapon system, about 15 percent of LCC is spent on development, about 25 percent on production and facilities, and 60 percent for initial logistics support and operating and support costs.

It is both interesting and noteworthy that this rough breakout of costs is shared not only by the commercial aircraft industry, as might be expected, but also by automobiles, televisions, and refrigerators.

The cost breakout in Figure 12-2 should in no way be construed to be the optimum categorization of LCC. It is important for the analyst to categorize costs in ways which will be meaningful to the decision maker. However, when LCC is used to compare alternative systems or designs, it is important to use the same basis for comparison. This implies that the same cost categories should be used in evaluating each system.

RELEVANT VERSUS INCREMENTAL COSTS

It should also be noted that the use of LCC in choosing among alternatives, only the relevant and incremental costs should be considered. Relevant costs are defined as expected future costs which differ under various alternatives.

To illustrate, suppose a cost analyst is attempting to find the cost impact of various maintenance concepts, i.e., three-level vs two-level, or in-house vs contractor operated, etc. Since air crew training costs are not affected by the choice of maintenance concept, these costs are not relevant and should not be considered in this particular analysis.

Similarly, incremental costs are defined as additional costs required to achieve a stated capability or level of activity. Incremental costing is very useful in analyzing replacement systems. In such cases, only the costs which differ from the old system need to be considered when evaluating the replacement system. Positive incremental costs are those required costs which exceed the costs already being incurred with the current system. On the other hand, cost savings to be achieved with the replacement system can be viewed as negative incremental costs.

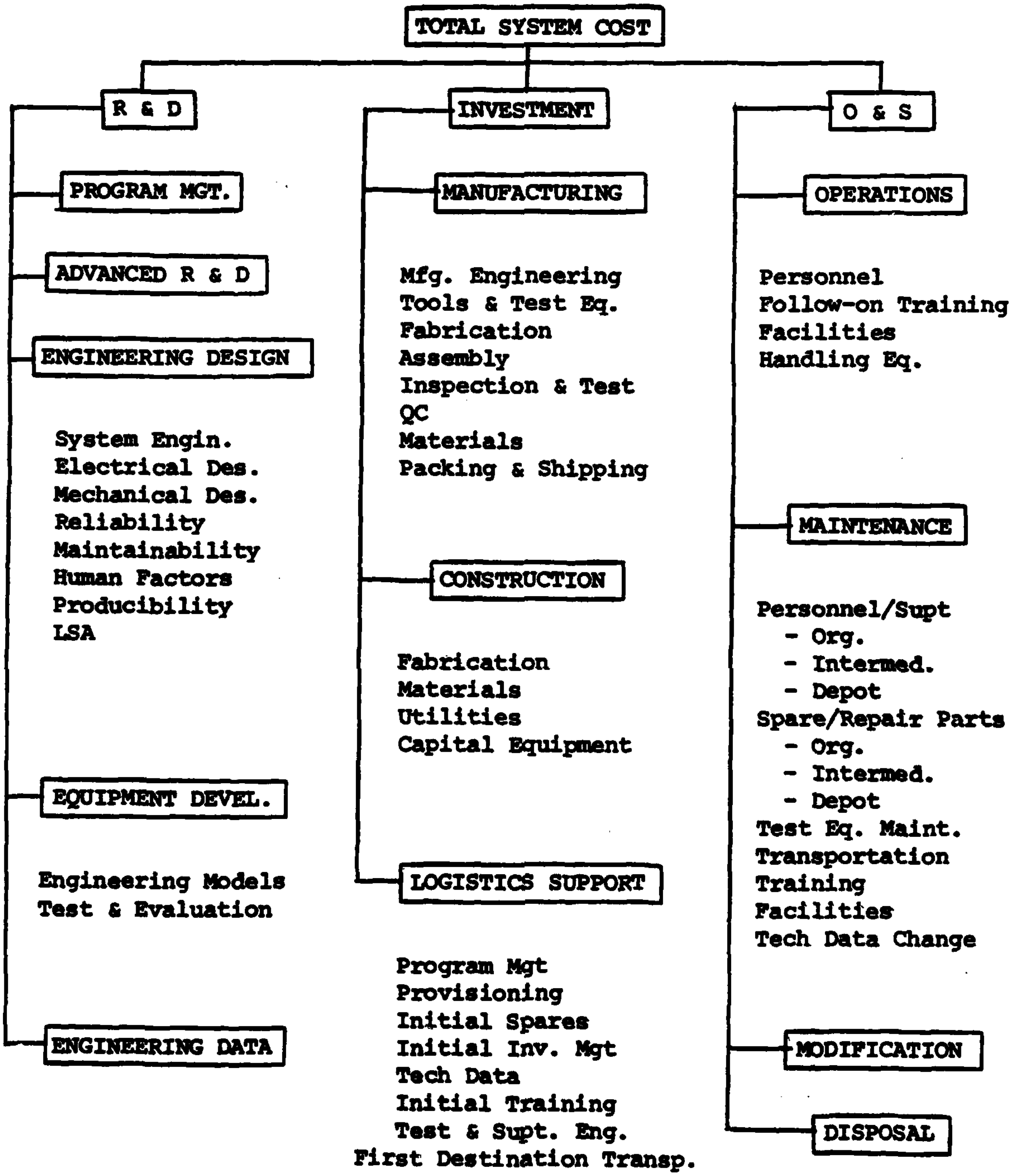
In the final analysis, the use of the concepts of relevant and incremental costs forces the decision maker to focus attention on the most important factors affecting the costs of alternatives.

LCC Models

Life cycle costing often involves extensive use of models. A LCC model is merely a representation of the significant cost relationships of the item/system under consideration. Models have been developed to provide estimates of total cost and estimates of design-related cost.

There are literally hundreds of life cycle cost models which exist today. Each has been constructed for a specific purpose/program. To be

Figure 12-2
LCC BREAKDOWN STRUCTURE



SOURCE: B. S. Blanchard, Logistics Engineering and Management

useful, however, each model should exhibit the following characteristics:

1. <u>Completeness</u>: The model should include all the elements of cost which are appropriate for the issue being considered. If two alternatives are under evaluation but replenishment spares cost the same for each system, then the model need not consider replenishment spares.

2. <u>Validity</u>: The model should represent reality. The analyst must verify the input data to insure that the model output represents the real world environment. Without model validity, the user would not be able to trust the output. In the validation process, the analyst must pay particular attention to logic and consistency.

3. <u>Sensitivity</u>: Each model should be sensitive to changes in parameters or design. It should be able to show the differences in costs which would occur when system configurations are altered. Each model should also be able to identify the major cost drivers of a system or subsystem. Management emphasis can then be concentrated on those parameters/items which are responsible for the greatest costs.

4. <u>Availability of input data</u>: Accurate input data must be available not only to constrict the model, but also for meaningful application of the model. Data availability has generally not been a problem. The problem lies in the accuracy of the data and the format in which it is collected. Both cause the analyst extreme difficulties. A problem also exists in the proliferation of data. For example, one O&S cost model used more than 200 data sources from both government and industry sources.

TYPES OF LCC MODELS

Life cycle cost models have been classified into categories which describe how the models are used or how they are formulated. The most common classifications are enumerated below. These classes are not mutually exclusive, i.e., a given model may be included under two or more categories.

There is a common misconception that a life cycle cost model contains the totality of life cycle costs for a system/component. This is generally not true. The vast majority of LCC models consider only a portion of total LCC. This is in keeping with the established concepts that only relevant and incremental costs should be considered. At the same time, however, there is occasionally a need to determine or calculate the total LCC for a system. Several models are available for accomplishing this task.

In general, each LCC model can be classified under one or more of the following categories:

1. <u>Accounting models</u>. Accounting models are cash register-like models in which the components of cost are aggregated through sets of

equations. Operating and support cost models are usually found in this category which accumulates costs in "accounts," such as depot maintenance, training, initial spares, etc. The Logistic Support Cost (LSC) model is basically an accounting model. The LSC model aggregates costs for spares, fuel consumption, personnel training, data, etc. The model also shows the percent of LCC encompassed by each cost category.

2. Cost factor models. These models estimate each cost element using a factor derived from Air Force experience on similar systems. A large collection of cost factors is maintained on a current basis in AFP 173-13, USAF Cost and Planning Factors. Among the thousands of factors contained in AFP 173-13 are the standard pay rates for all AF grades, fuel and oil consumption and costs per flying hour for each M/D/S aircraft, annual and per flying hour costs for depot maintenance, etc. Cost factor models have been used by the Air Force in estimating missile and aircraft squadron annual operating and support costs. Examples are the the Cost Analysis Cost Estimating (CACE) model and the Planning, Programming, Budgeting Annual Cost Estimating (BACE) model. Both models can be found in AFP 173-13.

A typical cost-factor equation contained in the BACE model for estimating depot maintenance cost is:

$$\text{Cost} = (\text{UE}) \cdot \left[(\text{UE Factor}) + (\text{FH Factor} \cdot \text{FH})\right]$$

where,

UE = number of aircraft (unit equipment)

FH = number of flying hours

3. Cost estimating relationship models. CER models are a form of parametric cost estimating which relates historical cost data to design, performance, or environmental parameters. CER models are statistically derived through regression analysis and have the advantage of using relatively few data inputs to estimate a large portion of total costs.

To illustrate, a major aerospace corporation formerly was using 15,000 lines of data to define the maintenance work load for a missile system. The company found that through the use of CERs only 108 lines of data were needed to define 91 percent of the work load.

The Modular Life Cycle Cost Model (MLCCM) for Advanced Aircraft is a parametric LCC model which uses a large number of CERs to estimate LCC. The model was developed for the Air Force Flight Dynamics Laboratory to conduct design/cost performance trade studies of advanced aircraft systems during conceptual and preliminary design phases. MLCCM has been loaded onto the AFLC CREATE computer system to provide maximum flexibility for users of the model.

Most CER models consist of many individual CER equations, each estimating a different component of LCC. A CER model for an aircraft system would have several CERs related to engine costs. For example, one CER could relate turbine intake temperature and thrust to engine manufacturing costs. Another might relate hours between removals to depot maintenance costs. By summing the individual CERs, an estimate of total LCC can be obtained by the analyst. And since the CERs are statistically derived, confidence intervals for the estimates can be constructed.

However, a problem arises when constructing a confidence interval for the total LCC which has been found by summing the CERs. This problem can be resolved by using a special technique which is explained in the final section of this chapter.

4. Economic analysis models. Economic analysis models are usually developed to assess costs versus benefits problems or to aid in lease versus buy decisions. These models enable the user to make trade-offs between investment and O&S costs and consider the time value of money through the use of present value discounting techniques.

The Army Corps of Engineers has frequently used economic analysis models to assess the benefits/costs of waterway projects such as dams and bridges. Recent developments have resulted in economic analysis models for use in determining the value of research projects. A model which relates research costs and potential benefits can be very useful in deciding whether a particular research project should be undertaken.

5. Simulation models. Used primarily in the logistic support cost area, simulation models are well suited in analyzing the cost sensitivity relative to major program changes, such as flying hours, maintenance concept, or operational scenario. The Logistics Composite (LCOM) model is an example of a simulation model widely used by the Tactical Air Command to assess the costs of various deployment and contingency plans.

6. Other models. A number of other LCC models have been developed to estimate costs associated with warranties, reliability improvement, manpower, inventory management, and level of repair (maintenance concept).

USES OF LCC MODELS

LCC models have been established to accomplish a wide variety of cost estimating tasks. Many of these tasks have already been described in Chapter 11, Cost Estimation. Although most models have been created for a specific purpose, such as to estimate production costs for a particular aircraft system, others can be used to estimate costs for a variety of aircraft systems. The Modular Life Cycle Cost Model for Advanced Aircraft falls into the latter category. The user should beware of the purpose for which a particular model was created. Each model cannot and should not be used indiscriminately. In other words, the peculiar requirements for each LCC estimating problem should be carefully assessed to ensure that the model selected is appropriate.

The list of uses and applications of LCC models is a long one. In addition to estimating all or a portion of total system LCC, LCC models have found wide application in break-even and sensitivity analysis, as well as cost effectiveness and benefit-cost analysis mentioned in a previous chapter.

LCC models can be used to assess the value of proposed engineering changes. Other models can and have been used to select an appropriate maintenance concept for a new system.

An important use of LCC models has been to establish cost goals, not only for acquisition costs, but O&S costs as well. A common technique is to establish an O&S cost target and reward/penalize the successful bidder when measured O&S costs fall below/above the targeted costs. This technique was used in a limited sense in the acquisition of the F-16.

In the final analysis, the concept of LCC is rapidly becoming a way of life within the DOD. Cost, in particular life cycle cost, is and will continue to be a predominant factor in the acquisition and maintenance of the force structure of the U.S. military establishment.

PREDICTION INTERVALS

In a previous section, it was noted that certain problems arise when an attempt is made to sum individual CERs. When CERs are developed statistically, confidence intervals (called prediction intervals) can be computed for each CER. Although a total cost estimate can be found by summing the costs estimated by each CER, a problem arises in the case where a prediction interval is desired for the total cost estimate. A solution to this problem was presented by J. A. Die Rossi.* The proposed solution encompasses two separate cases. The first is for summed totals with equal variances, i.e., the standard error for each CER is equal. The second case arises when the standard errors are not equal. In studying the solution methodology, it is readily apparent that the solutions closely resemble Cases 3 and 4 of the t-tests presented in Chapter 8 on hypothesis testing.

Users should refer to RM-5806 PR for the theoretical basis of the solution and its application. For direct application, a solution model is available on the AFLC CREATE computer system.

*J. A. Die Rossi, "Prediction Intervals for Summed Totals," Rand Memorandum RM-5806-PR (Santa Monica CA: The Rand Corporation, October 1968).

CHAPTER 13

REGRESSION ANALYSIS

Regression analysis is a technique for forecasting the value of some dependent variable based on the known or easily predicted value of at least one other variable (independent variable or variables) and making an assessment of the range of uncertainty in the value of the dependent variable.

Regression analysis is useless if the independent variable is as hard to predict as the dependent variable. A way to predict snow removal cost based on the amount of snowfall could be accurate to the penny and still not helpful--it is as hard to predict the amount of snowfall in advance as it is to predict costs.

There are three skills involved in regression analysis:

a) The skill of determining what easily predictable (or known) independent variable or variables cause the dependent variable to vary.

b) The skill of determining the relationship between the dependent and independent variables.

c) The skill of determining whether or not the relationship (or relationships) found are meaningful.

The technique should be:

1. Decide on the cause and effect relation.

2. Gather the data, each observation or case to contain a value for each variable.

3. Graph the data as a preliminary test of the hypothesis stated in step one.

4. Fit the equation.

5. Evaluate the results.

It is essential that these steps be followed. It is unforgivable to collect a set of data and then find what "fits best."

The cause and effect relation comes from knowledge, there is no way to detect it from the data. This implies one must be familiar with the item under consideration. If the analyst lacks this personal knowledge, expert

assistance must be sought. Engineers can sometimes suggest relationships. There are people at the AFPRO (if one exists) that may know. If there is a SPO connected with the items, someone there may be able to suggest cause and effect relations. Contractor personnel may be willing to provide information but the analyst must remember contractor personnel are not disinterested and their interest is not necessarily the same as that of the Air Force. Gathering the data may be very difficult indeed. Records get lost, or thrown away, data is not kept with analysis in mind. Events occur over time and time brings changes.

Unless the independent variable is time, one must make the data comparable. If costs are collected in dollars, it probably will be necessary to use some method of correcting for time, converting the "current year" dollars to "constant dollars" with the use of an appropriate index series. If the independent variable is time, regression does not apply; use a time series analysis.

In many cases it is necessary to correct for cost-quantity effects (see the Learning Curve chapter). Graphing the data is easy if there is only one explanatory (independent) variable, use regular graph paper or, if nonlinearity is obvious, use semi-log paper or log-log paper. Each type of paper gives a characteristic formula (equation) if the data appears to be a straight line.

Fitting the equation is a simple matter if you have a computer available.

- At AFLC bases the pricing function has a computer system--Copper Impact--available, several stored programs are available. The people at the pricing shop will probably help you use it.

- There are many other computer systems but access may be difficult--check with the Data Automation people on your base.

- For two variable problems some electronic calculators--HP65, HP38E, TI58, TI59 or equivalent are either wired to, or can be programmed to fit a regression line.

Evaluation of the results is not easy. There are sets of statistics that can be used for evaluation. These statistical tests assume there is no relation between the independent variable or variables and the dependent variable (The null hypothesis or H_o). Say you are only willing to take a 5% chance of being wrong. Then certain statistics are calculated, if the calculated statistic is greater than some critical value (determined by the chance you are willing to take) we say that could have happened only five times in one hundred (probability .05) if there was no relationship. But since it did happen, we reject the null hypothesis. Notice that five percent of the time we expect to be wrong, i.e., there will be no relationship among the variables but by chance we get a statistic greater than critical.

An example will show what we mean.

The following problem was extracted from a paper by Mr. Aubrey A. Yawitz (1). Mr. Yawitz does an excellent analysis but the only part we have copied is the problem and his cause and effect. The following analysis is all our own given the cause and effect.

The problem is to devise a cost estimating relationship to estimate labor hours in overhaul, the heavier the locomotive the more labor hours it takes to overhaul.

For our purposes we will hypothesize a linear relationship

$$\text{Labor hours} = a + b\ (\text{weight})$$

$$H_O:\ b = 0$$

H_O is the null hypothesis that will be tested after the equation has been fitted.

The data collected from a single depot in the years 1977 and 1978 are:

Observation	Locomotive Gross Weight (tons)	Unit Overhaul (Labor hours)
1	44	2208.00
2	60	2625.86
3	60	2425.33
4	60	2501.00
5	80	2361.50
6	100	3092.50
7	100	2913.75
8	100	3214.67
9	120	3481.00
10	120	3332.00
11	120	3003.50

The first step is to plot the data on a sheet of graph paper (see figure 13-1). The scales will be arbitrary but should have two characteristics:

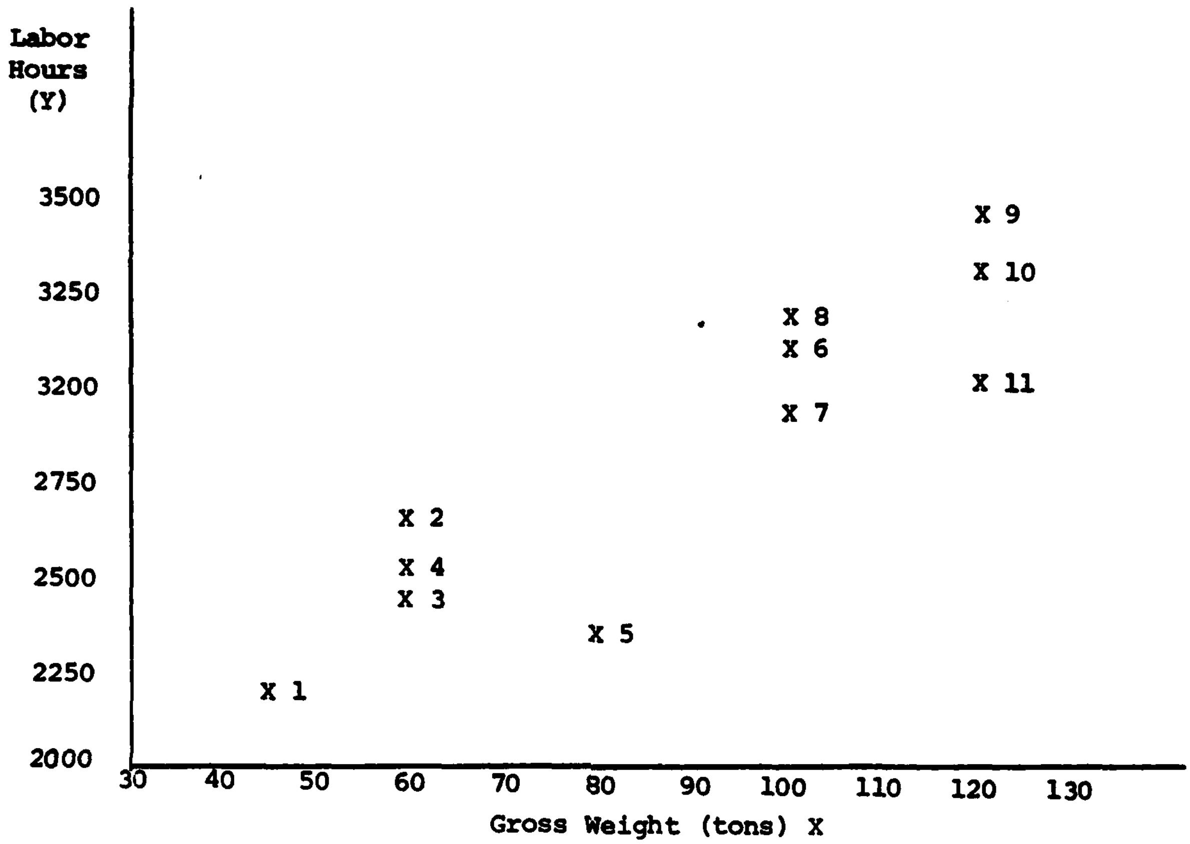
1. Wide enough so all the data can be plotted.

2. Scaled so reading is moderately easy.

The X axis is gross weight in tons, and the scale is 2 tons to the 1/10 inch or 20 tons for ONE LINEAR INCH of axis. Figure 13-1 shows a plot of this data. The Y axis is labor hours, the scale is 50 hours to the 1/10 inch or 500 HOURS FOR ONE LINEAR INCH of axis.

Plotting is the first, and rather rough test of the hypothesis of a linear relationship between weight as the causative variable and hours used in overhaul as the effect.

Figure 13-1. Data on Locomotives



To plot the first observation one reads weight equals 44 tons, hours equals 2208. On the weight axis read the 44 then up the Y axis to 2208, make a mark (I have used an X) and leave a number to identify the point. I used the observation number.

The rest of the points are plotted the same way.

When the data has been plotted it should be examined for outlines (data points well outside any pattern). The point for observation five appears to be such a point.

Outliers deserve to be investigated--there may be good reasons--a special type of unit, one unit designed to work in a special environment, a unit made under a causation system different from the rest--are examples. If such a cause is found it is reasonable to either adjust the data or exclude that data point from the study. Sometimes the outlier turns out to be simply a typographical error which can be corrected.

There is no statistical test that should enable an analyst to ignore a data point. The job is to test a hypothesis. If you exclude data that does not fit your hypothesis even before you do a formal test, there is no probability of rejecting that hypothesis.

In the AFIT/LS computer system there is a computer program to do regression. Usually the data is set up in a file in computer memory according to specific computer program instructions and then some regression program (there are several of them) is run according to instructions.

Nearly all computer programs have the same output, some optional, some standard. For this problem the analyst got the following outputs:

Equation: Labor hours = 1620.904 + 13.827 (WT)
Weight is in tons

Equation F score 38.824

Standard error of the b = 2.21908

t for b = 6.231

Arithmetic Mean of Y = 2832.646

Standard deviation of Y = 430.00

Variation in Y = 1849770.35

90% Confidence Level Prediction Intervals

X value	Lower Bound	Point Estimate	Upper Bound
0	1102.407	1620.904	2139.401
44	1813.028	2229.289	2645.550
60	2057.575	2450.520	2843.465
80	2349.257	2727.059	3104.860
100	2623.731	3003.598	3383.464
120	2881.265	3280.136	3679.008
160	3355.293	3833.214	4311.134

For those who do not have access to a computer the method of calculation of these numbers on a desk calculator is shown in the section on calculation at the end of this chapter.

Evaluation: One data point should be questioned; observation 5. But unless clear evidence of error is found, it should be kept in the data set. It was kept in the data set.

Statistical significance. There are three measures of statistical significance:

1. For the equations: An F score is used. Technically, F is a relation of variances, the "explained variance"[1] and the "unexplained variance"[2].

When the number value for F is determined it is compared to a standard from tables. F tables are divided into levels of significance, each of which has two arguments. One less than the number of parameters to be estimated in the head and the number of observations minus the number of parameters in the stub. In our case the arguments are 2-1 or 1 in the head and 11-2 or 9 in the stub.

If the calculated value of F is greater than the table value, one considers the result significant. For the case at hand we chose the 95% level of confidence with column 1, row 9 and got 5.12 from the table. This means that if there is no relation between cost and weight we would expect an F score of 5.12[3] or larger only one time in twenty trials. Since the F from the data is 38.824 we consider our results significant at the 95% level and reject the hypothesis of no relation between them.

2. The significance of the b of the equation. In two variable problems these two tests are equivalent, in larger problems they are not. In this case we use a calculated "t" statistic (parameter b divided by its own standard error) in our case 6.231 (13.827/2.219). This score is compared with a standard "t" table. The "t" table has two arguments, level of confidence in the head and number of observations less number of estimated parameters in the stub. In our case the stub value is 11 minus 2 or 9 at the 95% confidence level.

In the table column for 95% confidence level row 9 we find 2.262[4]. The "t" for our problem is 6.23. Since 6.231 is greater than 2.262 we reject the hypothesis that b = 0. This test is exactly analogous to the test for the equations, if b were indeed 0, i.e., there was no statistical relationship between cost and weight, only 5% of the time would a sample of 11 show a "t" as large or larger than 2.262. Remember, we expect to reject H_o some five percent of the time when it is true.

3. The third test of significance is not so precise as the others (it is the test of H: a = 0. There are very precise tests for this hypothesis but they are uncommon in canned computer programs. This much can be said if the prediction interval for X = 0 does not include Y = 0 the null hypothesis can not be rejected at the specified level of confidence. In our case the bounds for the 90% prediction interval when X = 0 are 1102.407 to 2139.401. This interval does not include zero hence the hypothesis a = 0 can be rejected at the 90% level of confidence. This is a conservative test of significance but whether or not a = 0 is not really very important.

Is regression worth the trouble? We know we could use univariate statistics, estimate the new cost as the average of the old costs with a

---

[1]Often called "mean regression sum of the squares"

[2]Often called "mean error sum of squares"

[3]Table from CRC Standard Math Tables, 16th Edition, The Chemical Rubber Co., Cleveland, OH, 1968, p. 588.

[4]Ibid, p. 583

statistical probability. If you use only the Y values as predictors you will always predict the net value of Y as the mean of the set within the interval.

Mean of Y + appropriate "t" [standard deviation of Y^2] + standard deviation of Y/sample size.

Say a 90% prediction interval for the next Y would be

$$2832.65 \pm 1.812 \sqrt{430.09^2 + (430.09^2/11)}$$

$$= 2832.65 \pm 1.812 \sqrt{201793.54}$$

$$= 2018.65 \leq Y \leq 3646.63$$

This says we expect Y to be between 2019 and 3647 but 5% of the time Y will be greater than 3647 and 5% of the time Y will be less than 2019. Our "best" guess all the time is 2832.65 or 2832.65 ± 28% of 2832.65.

If we use regression and the same confidence level the "best guess" depends upon the value of X, the weight of the locomotive in tons.

Our interval estimates would be:

for a 44 ton locomotive we say 2229.892 ± 18.67% of 2229.892

for a 60 ton locomotive we say 2450.520 ± 16.03% of 2450.520

for a 80 ton locomotive we say 2727.059 ± 13.85% of 2727.059

for a 100 ton locomotive we say 3003.598 ± 12.65% of 3003.598

for a 120 ton locomotive we say 3280.136 ± 12.16% of 3280.136

for a 160 ton locomotive we say 3833.214 ± 12.47% of 3833.214

For a different sample or data set the numbers would all be different. Our estimates are made from our specific example.

We have considerably improved our ability to predict the value of the manhours needed to overhaul a locomotive when we know the weight of that locomotive.

Instead of a single estimate the estimate depends upon the weight and none of the intervals are as wide as the 28% we would have with only costs instead of a cost estimating relationship.

One good question: Are those estimates good enough? The answer depends upon the needs. If the answer is yes, well and good. If no, then one must consider alternatives--are there other characteristics in addition to weight which are independent of weight and might cause costs? Is it possible to extend the sample? Are there other locomotives on which

overhaul data is available? Finally, if this CER is inadequate, what do you have that is better?

There is one very bad habit, try several variables on which one has data to find which gives the best fit. This is wrong.

Remember tests of significance! If there is no relation between variables the F score will be so large some specified percent of the time. Indiscriminant fitting is wrong because consecutive tests destroy probabilities. If one tests two separate models the probability of accepting at least one (assuming no relations among them) is about 10% instead of the 5%, or if you are using 90% level of confidence the probability of accepting at least one is nearly 20%. If one is depending upon statistical tests the rules of probability must be checked carefully.

SOLUTIONS WITH A DESK CALCULATOR

To calculate the numbers given in the example problem, construct a worksheet. Worksheets tend to be tedious and subject to errors. Careful work will help prevent errors but a check column is recommended.

Table 13-1 shows the worksheet; the steps in calculation are:

1. Copy the data into columns 1, 2, and 3

2. Sum column 2, get the sum of X = 964

3. Sum column 3, get the sum of Y = 31159.11

4. Fill in column 7, K = X + Y

5. Sum column 7, get the sum of K = 32123.11

6. Check your addition by checking:

   does the sum of K = sum of X + sum of Y?

   does 32123.11 = 964 + 31159.11?

It does, so we know the addition is correct.

7. Get the mean of X. Divide the sum of X (sum of column 2) by the number of observations (11), get mean of X equals

   964/11 = 87.6363... a continuing decimal the 6363 will go on forever

8. Get the mean of Y. Divide the sum of Y (sum of column 3) by the number of observations (11), get mean of Y equals

   31159.11/11 = 2832.6463...another continuing decimal

Table 13-1. Worksheet for Locomotive Data

Column No.	1	2	3	4	5	6	7	8
	OBS	X	Y	X^2	XY	Y^2	K	K^2
	1	44	2208.00				2252.00	
	2	60	2625.86				2685.86	
	3	60	2425.33				2485.33	
	4	60	2501.00				2561.00	
	5	80	2361.50				2441.50	
	6	100	3092.50				3192.50	
	7	100	2913.75				3013.75	
	8	100	3214.67				3314.67	
	9	120	3481.00				3601.00	
	10	120	3332.00				3452.00	
	11	120	3003.50				3123.50	
SUMS		964	31159.11	92336.00	2839275.40	90112509.37	32123.11	95883396.17
MEANS		87.6363	2832.6463				2920.282727	

The variation in X is 92336 - (964 times 87.63636363) or 7854.5454...

The variation in X and Y is 2839275.40 - (87.63636363 times 31159.11) or 108604.3250

The variation in Y is 90112509.37 - (31159.11 times 2832.646363...) or 1849769.750

The variation in K is 95883396.17 - (32123.11 times 2920.282727) or 2074832.90

9. Get the mean of K, divide the sum of K (sum of column 7) by 11

Mean of K = 32123.11/11 = 2920.282727...

10. Another check--does the mean of K equal the mean of X plus the mean of Y? Does 2920.282727 = 87.636363 + 2832.646364? It does so the means have been calculated correctly.

11. Accumulate the sum of the X squared in a calculator, this means square 44, square 60 and add to the square of 44 and so on until all eleven numbers have been squared and the sum accumulated. Resist the temptation to square individually, write down and then sum. This almost guarantees errors will creep in. Write the sum at the foot of the X^2 column, column 4.

12. Accumulate the sum of the cross products of X and Y in the calculator. This means take each observation, multiply the X and Y (for Obs. 1 use 44 times 2208.00) and accumulate. Write the sum at the foot of the XY column (col. 5).

13. Accumulate the sum of the squares of Y in the calculator and write the sum at the foot of the Y^2 column (col. 6).

14. Accumulate the sum of the squares of K (col. 7) in the calculator and write the sum at the foot of the K^2 column (col. 8).

15. Check the accuracy. If the sum of K^2 equals the sum of X^2 plus the sum of Y^2 + twice the sum of XY, the worksheet summations are correct.

Does 95883396.17 = 92336.00 + 90112509.37 + (2 times 2839275.40)

It does, so the major calculations are correct.

16. The variation in X is equal to the sum of the squares of X minus (the sum of X times the mean of X). Remember to do the multiplication inside the parenthesis, then subtract.

17. The covariation in X and Y is equal to the sum of the cross products (col. 5) minus (the mean of X times the sum of Y). Remember, multiply numbers in parenthesis, then subtract product.

18. The variation in Y is the sum of the squares of Y (col. 6) minus (the sum of Y times the mean of Y).

19. The variation in K is the sum of the squares of K (col. 8) minus (the sum of K times the mean of K).

20. The last check of the worksheet--does the variation in K equal the variation in X plus the variation in Y plus twice the covariation in X and Y?

Does 2074832.90 = 7854.5454 + 1849769.75 + (2 times 108604.325)?

To the second decimal place it does and the difference is due to rounding. The calculations on the work sheet are correct.

To find the regression equation: Y + a + bX + e the calculations are simple once the worksheet is complete.

b = covariation in X and Y divided by variation in X.

b = 108604.325 divided by 7854.5454

b = 13.8269396

and

a = mean of Y minus (b times mean of X)

a = 2832.646363 - (13.8269396 times 87.6363)

a = 2832.646363 - 1211.7427

a = 1620.90366

or in easier terms

Y = 1620.904 + 13.827X + e

In order to calculate the numbers we need for statistical inference we need a table of variation and variance. Table 13-2 shows the needed table.

The total variation is simply the variation in Y, thats all there is in the problem.

Degrees of freedom is a statistical idea, For total variance it's always the number of observations less one for regression problems.

The total variance is the variation divided by degrees of freedom.

The standard deviation of Y is the square root of the total variance.

The variation in Y that is "explained" by the variation in X is a statistical concept. We can avoid a lot of theory if we simply take it as

Explained variation = b times covariation in X and Y

In this case

Explained variation = 13.8269396 times 108604.325 (the covariation in X and Y was calculated on the worksheet)

Explained variation = 1501665.444

Table 13-2. Table of Variation and Variance for Locomotive Overhaul Problem

Source	Variation	Degrees of Freedom	Variance	
Total	1849769.750	10	184976.750	Standard deviation of Y = 4300.89
Explained	1501665.444	1	1501665.444	
Unexplained	348104.306	9	38678.256	Standard error of estimate = 196.66788

F = Explained variance divided by unexplained variance

= 1501665.444 / 38678.256 = 38.825

Standard error of b = standard error of estimate divided by square root of variation in X

= $196.66788 / \sqrt{7854.5454}$

= 196.66788 / 88.625873

= 2.21908

In two variable regression there is only 1 degree of freedom in the explained variation, hence explained variation and variance are the same (variance is always variation divided by degrees of freedom).

The unexplained variation, that is the variation of Y not explained by the variation in X is total variation minus explained variation. The degrees of freedom in two variable regression for the unexplained is always the number of observations less two, here 11-2 = 9.

The unexplained variance is unexplained variation divided by degrees of freedom, here 348104.306 divided by 9 = 38678.256 is the unexplained variance.

In statistics, the standard error of estimate is the square root of the unexplained variance,

$$\sqrt{38678.256} = 196.66788$$

This is our estimate of the standard deviation of the e in the equation Y = a + bX = e, and the e is distributed around that line $Y_C = a + bX$.

F score and standard error of b are calculated on the table of variation and variance, the only statistic we have to calculate is T score for b = b divided by its standard error or t for b = 13.8269396 divided by 2.21908 = 6.231.

In making estimates one uses the estimating equation and assigns the relevant X value. If one wishes to estimate the labor hour overhaul cost of a 44 ton locomotive one says:

$$Y_c = 1620.904 + 13.826936\ (44)$$

$$Y_c = 1620.904 + 608.385$$

$$Y_c = 2229.289$$

This is our point estimate of the value of Y. To construct a prediction interval we need a standard error of prediction at X = 44. This is

$$\text{Std error of estimate times} \sqrt{1 + \frac{1}{11} + \frac{(\text{Assigned values of X - mean of X})^2}{\text{Variation in X}}}$$

The standard error of estimate is from the table of variation and variance; the number of observations (11) is from the worksheet. Also from the worksheet we get the (a) variation in X, and (b) mean of X.

The assigned value of X is 44.

$$196.66788 \text{ times} \sqrt{1 + \frac{1}{11} + \frac{(44 - 87.636363)^2}{7854.5454}}$$

$$= 196.66788 \text{ times} \sqrt{1 + .090909091 + \frac{(-43.636363)^2}{7854.5454}}$$

$$= 196.66788\sqrt{1.0909091 + \frac{1904.132176}{7854.5454}}$$

$$= 196.66788\sqrt{1.0909091 + .2424242}$$

$$= 196.66788\sqrt{1.9333333...}$$

Standard error of prediction at X of 44

$$= 196.66788 \text{ times } 1.15470054$$

$$= 227.0925064$$

A prediction interval is always: a point estimate $\pm$ some appropriate t times a standard error of prediction. From a t table we find a 90% confidence level of t at 9 degrees of freedom is 1.8333 so our prediction interval is:

$$2229.289 \pm 1.833\ (227.0925064)$$

$$2229.289 \pm 416.261$$

Our lower bound is: 1813.0284

Estimate: 2229.289

Our upper bound is: 2645.550

For practice you may wish to calculate some of the other intervals shown in the text.

REFERENCES

Yawitz, Aubrey A., Cost Memorandum 79-8.

Depot Overhaul Costs for Diesel Electric Locomotives, February 1979. U.S. Army Troop Support and Aviation Materiel Readiness Command, Comptroller Cost Analysis Division, 4300 Goodfellow Blvd, St. Louis, MO 63120.

There are many excellent books on regression. Two of them are:

Draper, Norman R. and Harry Smith Jr. Applied Regression Analysis, 1966. John Wiley & Sons Inc. New York, N. Y.

Johnson, G. Econometric Methods, 2nd Ed, 1973, McGraw-Hill Book Company, New York, N. Y.

There are many good sets of tables for statistics; the one used here is:

Standard Math Tables, Chemical Rubber Co., Cleveland, OH. These tables are revised frequently and the current revision will be the only one available.

CHAPTER 14

THE LEARNING CURVE

There is considerable evidence that we learn from experience. A rather extensive body of theory, called learning curve theory, has grown around the idea.* This theory, although not without its detractors, has proven very useful in the area of cost estimating.

The learning curve gives us a way of estimating the direct costs of some items. The use of learning curve techniques will result in an "estimate," or an "expected" value that should be close "on the average." No calculations in this paper are meant to indicate that we can estimate costs accurately to 3 or 4 digits. We can use theoretical constructs to give a "best estimate" and that is all.

The three basic tools for use of learning curve theory are log-log graph paper, a good calculator, or access to a computer terminal.

a. One of the most useful tools for dealing with learning curves is graph paper with coordinates scaled so numbers can be plotted as logarithms. Both X and Y axes are so scaled. This type of graph paper is called log-log paper. It has the characteristic that equal ratio changes plot as a straight line. Say the three points (X_1Y_1), $(2X_1, .8Y_1)$, $(4X_1, .64Y_1)$ (each X is twice the previous X and each Y is .8 times the previous Y) these three points will plot in a straight line on log-log paper. (To illustrate this let $X_1 = 1$ and $Y_1 = 1000$, plot the points (1,1000), (2,800), (4,640), and see if they fall on a line.)

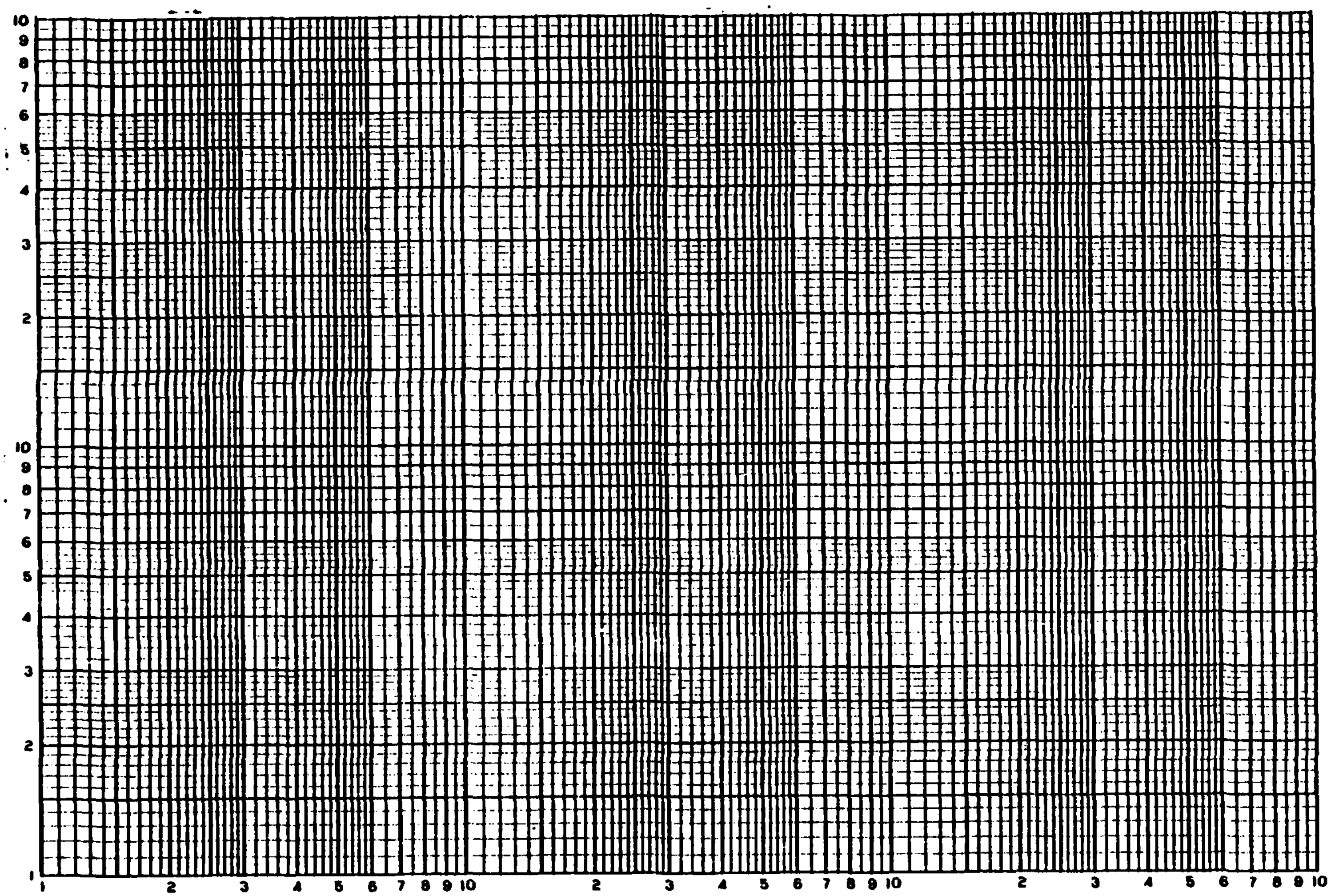
Figure 14-1 is a sheet of log-log paper. Note the scales are peculiarly spaced. There is no zero because zero does not have a logarithm, the distance from four to five is much less than the distance from one to two, in fact, the distance from five to ten is the same as the distance from one to two. The change in distance represents a change in ratio, from four to five is from X = 4 to 1.25X = 5. The distance from one to two is from X = 1 to 2X = 2, this latter is the same as the distance from X = 5 to 2X = 10, the same ratio as 1 to 2.

It is important to remember that equal distances on the paper represent equal changes in proportion--if one goes from X = 1 to 10X = 10 the result is a tenfold increase, so the second time this distance is covered one is going from X = 10 to 10X = 100, if there was a third cycle it would be from X = 100 to 10X = 1000. This scaling requires care in reading or in plotting.

*NOTE: Yelle, Louis E., "The Learning Curve Historical Review and Comprehensive Survey," Decision Sciences, Vol. 10, No. 2 (Apr 79) contains an extensive bibliography.

Figure 14-1

LOG-LOG Paper



The divisions of logarithmic scales also can cause trouble. Note the space from 1 to 3 and 10 to 30 have 20 minor divisions per major division. From 3 (or 30) to 6 or (60) there are 10 minor divisions per major division. From 6 to 10 (or 60 to 100) there are five minor divisions per major division.

Also, note that things tend to crowd up, measuring from 1 two minor divisions to 1.1 is about 4.5 milimeters on this paper. Starting at five and going right about 4.5 mm gets you to 5.5 or 5 times 1.1. If you start at 9 and go to the right about 4.5 mm you get to 9 times 1.1 or 9.9, you have nearly gone out of the 9 range. This explains the crowded conditions in the range nine to ten or 90 to 100.

This kind of paper can be secured from any university bookstore.

b. The second basic tool is a good calculator. One that has logarithms, and can handle powers and fractional exponents. While a good analyst can complete all the estimates using graph paper, a calculator allows more accuracy and provides a check on the graphical results. About the only thing you cannot easily do with a calculator is calculate the slope of the curve from actual data.

c. The third basic tool is access to a computer terminal. There is generally a learning curve program available which, given data, will do all the computations. This does not relieve you of the responsibility for graphing the data! A graph can highlight changes in the production process that would never be flagged otherwise. Once again, access to a terminal or even a fancy calculator are not required, but they can help.

Formulations:

The learning curve comes in two variants

Unit Curve (U.C.) $Y_X = AX^b$

where: Y_X is the cost of the Xth unit, A is a constant (the cost of unit one), X is the sequential unit number (usually limited to production units) and b is a constant, a negative decimal fraction, and

Cumulative Average (C.A.) $\overline{Y}_X = AX^b$

where: $\overline{Y}_X$ is the average cost for units 1 to X inclusive

A is the cost of unit one
X is the sequential unit number
b is a negative decimal fraction.

What is Being Estimated?

If Y_X or $\overline{Y}_X$ is in dollars it must be in constant dollars, but usually cost is expressed in real terms, for example:

Labor hours for manufacturing labor
Tooling hours for direct tooling
Engineering hours
Pounds of material

Slope:

One often hears about "slopes" of learning curves, the slope is the percent the Y is of the former Y as X doubles.

$$\text{Slope} = \frac{Y(2X)^b}{Y(X)^b} \ *100\%$$

If the "unit curve" $Y_X = AX^b$ is the learning curve and has a 80% slope, the points (X,Y), (2X, .8Y) are on the curve. For instance, if A = 1000 the following table describes an 80% curve.

Unit	Cost in Labor Hours
1	1000
2	800
4	640
8	512
16	409.6

The value of intermediate units can be determined by the formula.

$$Y_X = AX^{-.321928}$$

thus, unit 3 would cost

$$Y_3 = 1000(3)^{-.321928}$$
$$= 1000(.7021037)$$
$$Y_3 = 702.1 \text{ or unit 3 costs about 702.1 labor hours}$$

and unit 5 would cost

$$Y_5 = 1000(5)^{-.321928}$$
$$= 1000(.59563734)$$
$$Y_5 = 595.6 \text{ approximately}$$

Percent slopes are applied to cost as X doubles. If A = 1000, a 70% curve looks like this

Unit	Cost in Labor Hours
1	1000
2	700 (70% of 1000)
4	490 (70% of 700)
8	353 (70% of 490)
16	240.1 (70% of 343)

For <u>this</u> 70% curve the formula is

$$Y_X = 1000X^{-.514573}$$

Notice this formula differs from the previous one only in the exponent--that difference makes it a 70% curve not an 80% curve.

As with an 80% curve the value of units 3 and 5 could be approximated with the following results:

$$Y_3 = 568.2 \text{ approximately}$$

$$Y_5 = 436.8$$

One can use this type of calculation to estimate direct costs of a particular unit, the total cost of a specified number of units or the average cost of some specified number of sequential units.

<u>Example of Unit Curve</u>:

A single example of results from an 85% curve

$$\text{Unit Curve } Y_X = 1500X^{-.234465}$$

Table 14-1

Unit No. (X)	Cost of Unit (Y_X)	Cost of X Units (TC_X)	Average Cost For X Units (TC_X/X)
1	1500.0	1500.0	1500.0
2	1275.0	2775.0	1387.5
3	1159.4	3934.4	1311.5
4	1083.8	5018.2	1254.5
5	1028.5	6046.7	1209.3
6	985.5	7032.2	1172.0
7	950.5	7982.6	1140.4
8	921.2	8903.8	1113.0
9	896.1	9799.9	1088.9
10	874.2	10674.2	1067.4

NOTE: Detail may not add because of rounding.

This is a mathematical construct called the unit curve or Boeing Formulation of the learning curve. Actual data would be subject to many random influences and would be somewhat different.

If you need practice using log-log paper, plot the data above. (At least the Cost of Unit X and Average Cost for X Unit.) Figure 14-2 shows the results of this plot. Note that the unit line, the one that shows the cost of unit X, is straight on log-log paper, but the plot of the averages is not a straight line.

Example of Cumulative Average Curve:

There is a different form of the learning curve. The Cumulative Average Curve.

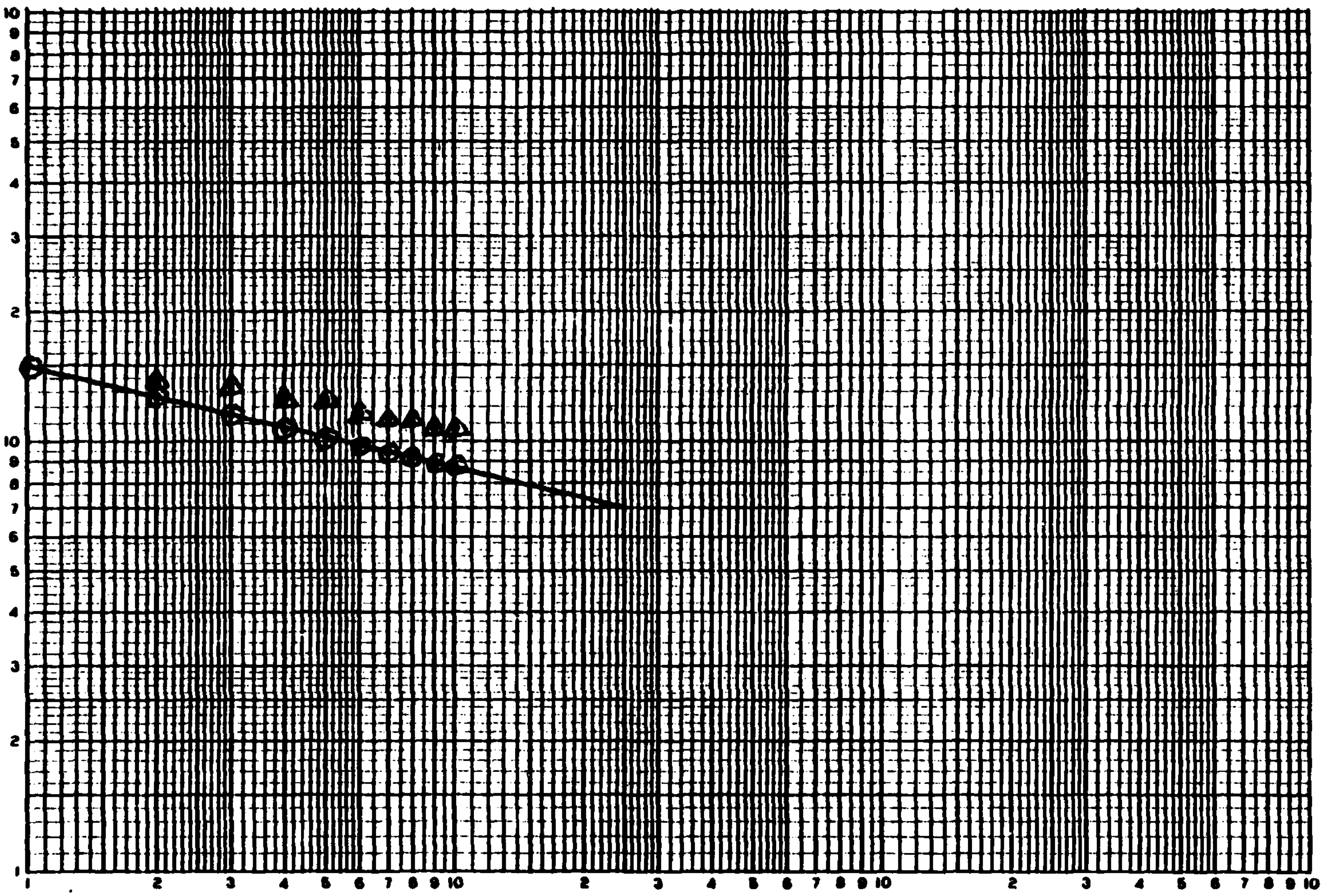
$$\bar{Y}_X = AX^b$$

The mathematical function is identical but it has a different meaning. This curve is increasing in popularity. In 1970, 95% of the curves studied by Defense Contract Audit Agency were unit type. Today many more Cumulative Average Curves would be found but no recent study is available.

Using the values of A and b as before A = 1500, b = -.234465, an 85% cumulative average curve would give results like Table 14-2.

Figure 14-2

Plot of 85% Unit Learning Curve
△ Average Cost for X Units
⊙ Cost of Unit X



85% Cumulative Average Curve $\bar{Y}_X = 1500X^{-.234465}$

Table 14-2

Unit No. (X)	Average Cost For X Units ($\bar{Y}_X$)	Total Cost of X Units ($TC_X = X\bar{Y}_X$)	Cost of Unit X ($Y_X = TC_X - TC_{(X-1)}$)
1	1500.0	1500.0	1500.0
2	1275.0	2550.0	1050.0
3	1159.4	3478.1	928.1
4	1083.8	4335.0	856.9
5	1028.5	5142.5	807.5
6	985.5	5912.8	770.3
7	950.5	6653.4	740.6
8	921.2	7369.5	716.1
9	896.1	8064.9	695.4
10	874.2	8742.3	677.4

NOTE: The numbers in column two of Table 14-2 are identical to column one of Table 14-1 but this time instead of being the cost of unit X they are the average cost for units 1 to X inclusive.

The average is cost per unit; therefore, if one multiplies the average by the number of units, the product will represent the total cost of that many units, or as the table says, the total cost of X units (units 1 through X inclusive).

$$TC_X = X\bar{Y}_X = XAX^b = AX^{b+1}$$

The last column is the cost of unit X which is, of course, the difference between the total cost of X units and the total cost of (X-1) units. Using the X = 5 row as an example.

$$\bar{Y}_5 = 1500(5^{-.234465}) = 1028.5$$

The cost of unit five is:

$$Y_5 = (5)(1028.5) - (4)(1083.75)^*$$

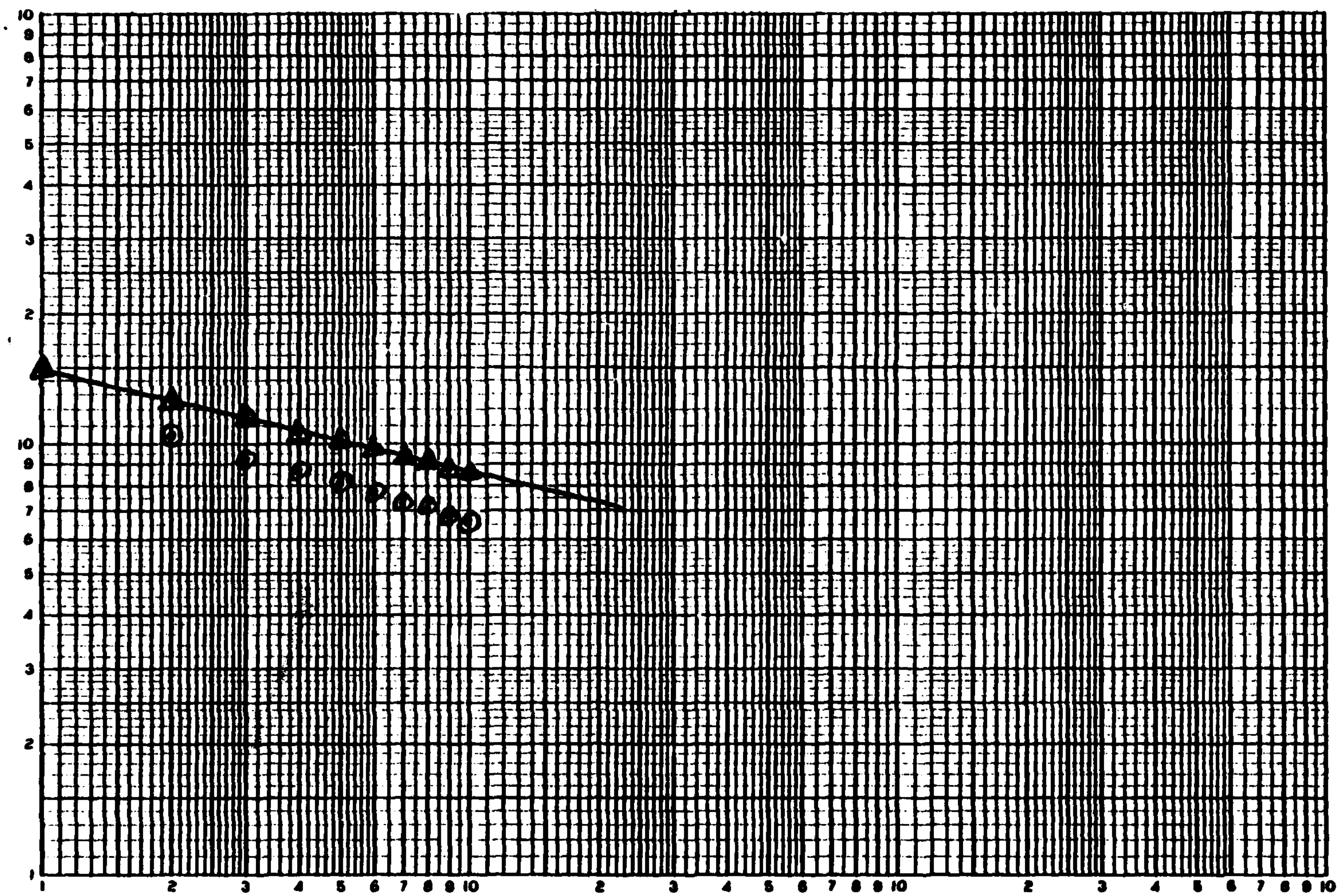
$$= 5142.5 - 4335.0 = 807.5$$

*Rounding makes this operation appear incorrect.

This data is plotted at Figure 14-3. Notice the cumulative average curve is a straight line while the unit curve is not a straight line.

Figure 14-3

Plot of 85%
Cumulative Average Curve
△ Average Cost for X Units
⊙ Cost of Unit X



Choosing a Formulation:

The coice of which form of the learning curve to use, unit or cumulative average, is not as difficult as it would appear. Perhaps the analyst knows the contractor who furnished the data and either the analyst or someone familiar with the company knows which form that company uses. If this knowledge is lacking, the analyst simply has not done the preparation needed for a cost estimate.

Be Sure LC Theory Applies:

There is a strong probability the learning curve is misapplied on occasion. There are instances where the learning curve theory does not apply. For example:

a. If the Air Force is buying commercial items from a contractor who is also making the item for sale to the public, the learning curve makes no sense. Even if one knew the company had a learning curve, how does the AF know which number items it bought?

b. Many analysts are convinced there is no learning curve for mass-produced items (those whose production is at or above 60,000 units annually).

c. Some accounting systems would not reveal a learning curve even if it existed. The comptroller of one of the "big-three" automobile companies testified under oath his company had no way to detect the cost of a sedan as it came off the assembly line. The best the company could do was know how much it cost to run that line for a period of time and by knowing how many cars were made during that period they could furnish the average cost of the cars made during that period.

c. One author claims learning curves are due to undetected nonrecurring costs in lot one. An example will be given later.

Given a learning curve problem, what should an analyst do? As a first step he should plot the data that purports to be a learning curve, plot both the unit values and the cumulative average values.

Plotting Lot Data.

The plotting done above is easy, the values for each unit were given, this is rarely the case. One usually has data on a lot (direct cost data) and it is necessary to edit the data in order to plot it. One needs a point, a specific X and a specific Y value to plot. The data, however, looks like that in Table 14-3. The data in Table 14-3 is hypothetical; it follows a perfect 83% learning curve, unit type, with a unit one value of 2500.

Table 14-3

Data for Plotting Learning Curve

Lot No.	Lot Size	Lot Value (Labor Hours)
1	10	16967
2	15	17424
3	20	19242
4	20	17017
5	40	30341

If this data is to be plotted it is necessary to find some way to represent each lot with one (or more) points on a sheet of log-log paper. The solution is to find a single point to represent the data for the unit curve and another point to represent the same lot for the cumulative average curve.

To accomplish this job in an orderly way it is advisable to construct a work-sheet, a guide for the process. Worksheets are, of course, personal and each analyst feels free to make their own, but one such as Figure 14-4 is recommended. If you choose this form leave room on the right for three more columns. We are going to get the points for the unit curve first then later the points for the cumulative average.

The start of the editing process is shown in Figure 14-4a. The information available has been entered and a zero placed in the cumulative units column.

The columns are:

Lot Number: Simply a way of identifying the lot.

Lot Size: The number of units in the lot.

Lot Value: The total cost of the lot in units specified, in this case labor hours.

Cumulative Units: The total number of units that were, or will be, produced when this lot is completed. This column represents a summation of the lot size through the lot being entered.

Lot Midpoint: A convenient method of identifying the middle of the lot. Fractional units are permissable in this column.

Figure 14-4. Partial Worksheet for Estimating Plot Points

Figure 14-4a. Layout Start of Edit

Lot No.	Lot Size	Lot Value (Labor Hours)	Cumulative Units	Lot Mid-Point	Lot Plot Points Unit Curve PP_X	PP_Y
			0			
1	10	16967				
2	15	17424				
3	20	19242				
4	20	17017				
5	40	30341				

Figure 14-4b. Worksheet completed for Unit Curve

Lot No.	Lot Size	Lot Value (Labor Hours)	Cumulative Units	Lot Mid-Point	Lot Plot Points Unit Curve PP_X	PP_Y*
			0			
1	10	16967	10	3.33	3.33	1697
2	15	17424	25	7.5	17.5	1162
3	20	19242	45	10	35.	962
4	20	17017	65	10	55.	851
5	40	30341	105	20	85.0	759

*Three digit accuracy is enough for plotting

PP_X is the X plot point for the lot--the midpoint of the lot plus cumulative units through the previous lot, fractions permissable in this column.

PP_Y: The Y plot point, the total value of the lot divided by number of units in lot (lot value) ÷ (lot size), the average value for the lot.

A zero was entered in the cumulative units because there has been no previous production, if this work sheet were a continuation, total previous production would be entered here--an uncommon occurrence.

I prefer to fill in cumulative units for each lot as the last entry for the lot.

When the work sheet is prepared, entries calculated are usually entered only to accuracy one can plot.

Step 1 Find lot mid-point for the first lot. If the lot size is less than ten, the lot midpoint is the lot size divided by 2. This is not true for the example. If the lot size is 10 or more, as here, the lot midpoint is the lot size ÷ 3, 10 ÷ 3 = 3.33, enter 3.33 in lot midpoint column.

Step 2 Find PP_X for first lot, PP_X = lot midpoint, enter 3.33 in PP_X for lot 1.

Step 3 Find PP_Y for lot, PP_Y = lot value ÷ lot size 16967/10 = 1697 (few people can plot even this accurately).

Step 4 Fill in cumulative units column, in this case no previous production so at conclusion of lot one, ten units will have been produced, enter a 10.

The following steps will be repeated as often as necessary to complete work sheet:

Step 5 Get lot midpoint for lot.
Lot midpoint = lot size/2, enter on worksheet

Step 6 Get PP_X for lot.
PP_X = lot midpoint plus all previous productions.
= lot midpoint plus cumulative units through previous lot.

Step 7 Get PP_Y for lot.
PP_Y = lot value/lot size

Step 8 Get cumulative units.
Cumulative units = cumulative units through previous lot plus lot size for this lot.

If not the last lot go back to Step 5 and repeat.

Each step should be checked with results in Figure 14-4b.

This part of the plot should always be carried out even if one suspects the cumulative average curve is the true formulation.

Figure 14-5 shows how this data plots on log-log paper.

This plot shows unit one costs of 2500 on an 83% slope.

The 2500 comes from reading the curve at X = 1. The 83% or slope of the curve is measured as follows:

Select an X and 2X-, in this case X = 20, 2X = 40 was chosen.

Read the curve at the X values, the points are (20, 1115), (40, 925). Divide Y_{2X} by Y_X and change to percent, 925/1115 = .83 = 83%.

If in doubt, select several different X, 2X pairs, and average the results.

Note that four of the points are perfect (or nearly perfect) fit but lot one is too far to the left. This is characteristic of first lots, they are frequently ignored while fitting, then a check is made.

The b of a learning curve can be calculated

$$b = \frac{\log \text{"slope"}}{\log 2}$$

where "slope" means the decimal fraction not the percent. In this case

$$b = \frac{\log (.83)}{\log (2)} = \frac{-.08092191}{+.30103000} = -.26881676$$

We know: $1696.7 = 2500X^{-.26881676}$

so $.678680 = X^{-.26881676}$

and $\log (.678680) = -.26881676 (\log (X))$

so $-.16833495/-.26881676 = \log X$

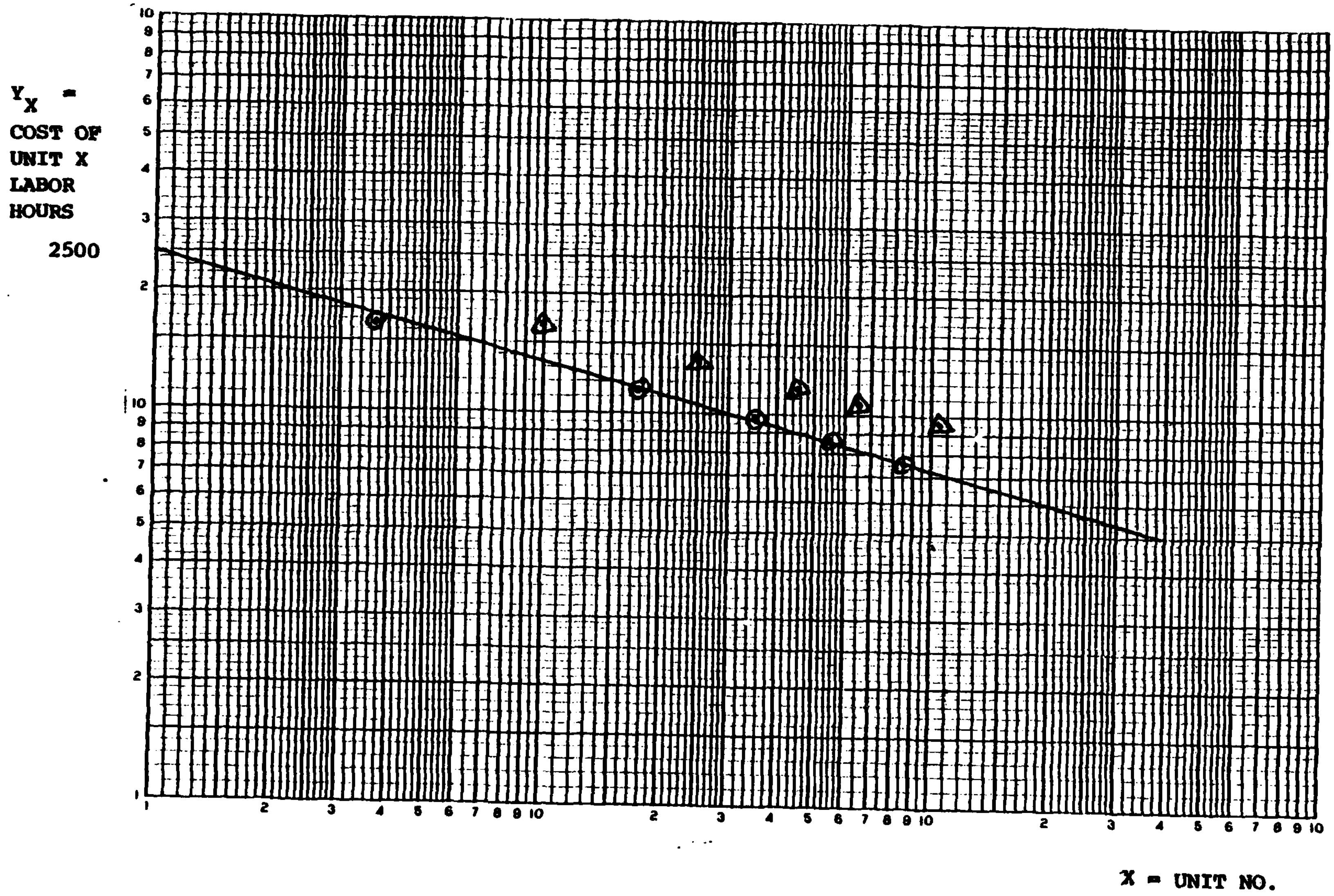
$= .62620704 = \log X$

$X = 4.229$

This means our rule of thumb X = 3.33 is about .9 units too low. If we moved the first lot plot point to (4.229, 1697) it would be on the line (or extremely close) so we can feel our curve is "right."

Figure 14-5

Unit Curve
Cumulative Average
⊙ Plot of Unit Curve
△ Plot of Cumulative Average Points



Most practioners of learning curve analysis do not check the first lot point, if its to the left of the line they feel satisfied.

Now let us set up our worksheet to plot the cumulative average curve.

Our worksheet needs three more columns, a cumulative total cost column, and PP_{XE}, $PP_{\bar{Y}}$ for plot points. Enter a zero above the row for the first lot in the cumulative total cost.

The cumulative total cost column is the sum of lot values through that lot. Figure 14-6a shows these added columns.

The plot point X for each lot is the number of units that will have been made at the end of the lot, this is simply the cumulative units through that lot.

The plot point Y is the cumulative average for all units made by end of lot,

$$PP_{\bar{Y}} = \text{cumulative total cost} \div \text{cumulative units}$$

Having gone through the unit curve plot points, the steps for calculating the cumulative average plot points are:

Step 1 Add the lot value to the cumulative total cost column (If its the first lot, simply enter lot value in cumulative total cost column).

Step 2 Enter cumulative units through the lot in the PP_{XE} column (The X plot point is number of units at the <u>end</u> of the lot).

Step 3 Divide cumulative total cost by cumulative units and enter quotient in the $PP_{\bar{Y}}$ column. (As the name indicates the cumulative average is average cost for units 1 through X.)

Step 4 If its not the last lot, go back to step one and repeat the sequence.

The completed worksheet is shown in Figure 14-6b.

This data is plotted in Figure 14-5. The plot behaves as it should, the result in nonlinear (on log-log paper) and the cumulative average curve would be above the unit curve.

The data above is all part of learning curve and plots as it should. Now, consider the data in Table 14-4.

Figure 14-6

Figure 14-6a Layout for Start of Edit for Cumulative Average

Lot No.	Lot Size	Lot Value (Labor Hours)	Cumu-lative Units	Lot Mid-Point	Lot Plot Point Unit Curve PP_X	PP_Y	Cumulative Total Cost	Plot Points Cum Average Curve PP_{XE}	$PP_{\overline{Y}}$
			0				0		
1	10	16967	10	3.33	3.33	1697			
2	15	17424	25	7.5	17.5	1162			
3	20	19242	45	10	35	962			
4	20	17017	65	10	55	851			
5	40	30541	105	20	85	759			

Figure 14-6b Completed Worksheet

Lot No.	Lot Size	Lot Value (Labor Hours)	Cumu-lative Units	Lot Mid Point	Lot Plot Points Unit Curve PP_X	PP_Y	Cumulative Total Cost	Plot Points Cum Average Curve PP_{XE}	$PP_{\overline{Y}}$
1	10	16967	10	3.33	3.33	1697	16967	10	1697
2	15	17424	25	7.5	17.5	1162	34391	25	1376
3	20	19242	45	10	35	962	53633	45	1192
4	20	17017	65	10	55	851	70650	65	1087
5	40	30341	105	20	85	759	100991	105	962

Table 14-4

Lot No.	Lot Size	Cumulative Units	Cumulative Average (labor hours)
1	10	10	3857.8
2	15	25	2057.8
3	25	50	1457.8
4	20	70	1286.4
5	40	110	1130.6

If one uses the cumulative average theory and plots this data and draws the "best" line one gets a cumulative average curve, unit one cost about 11,000, slope about 70%. Since the points do not fall on the line, there is some room for judgment.

Example of a Problem with Nonrecurring Costs:

Figure 14-7 shows this plot and learning curve. The plot looks reasonable, and one could get the impression this curve described the output. This is a false impression. The whole picture is caused by a nonrecurring cost in the first lot. Let us reconstruct the data so we can plot the unit curve.

Lot 1: There were 10 units in this lot, they average 3857.8 hours, the cost of the lot is 10 (3857.8) = 38578

of course, the average cost is 3857.8.

Lot 2: There were 15 units in this lot. At 25 units the cumulative average is 2057.8. The cost of this lot is then:

Cumulative total for 25 units = 25(2057.8) =	51445
Minus cumulative total for 10 units	38578
Cost of lot	12867
Average cost for lot	857.8

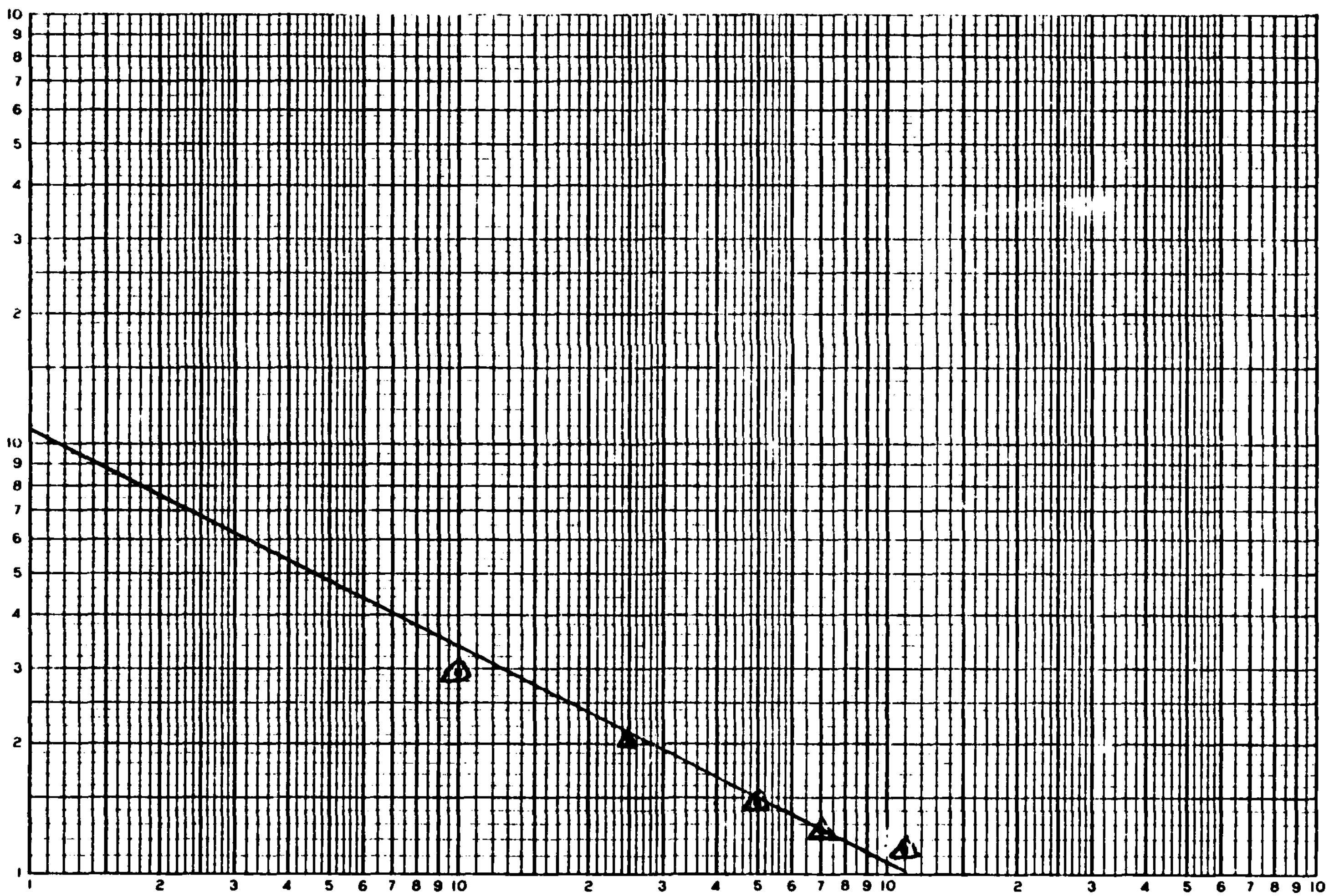
Lot 3: There were 25 units in this lot. At 50 units the cumulative average is 1457.8. The cost of this lot is then:

Cumulative total for 50 units = 50(1457.8) =	72890
Minus cumulative total for 25 units	51445
Cost of this lot of 25 units	21445
Average cost for this lot	857.8

Lot 4: There were 20 units in this lot, at the end of the lot 70 units had been made. Cumulative average is 1286.4. Cost of this lot is:

Figure 14-7

Cumulative Average Plot and Curve
for Data in Table 14-4



Cumulative total for 70 units = 70(1286.4) =	90048
Minus cumulative total for 50 units	72890
Cost of this lot	17158
Average cost for this lot	857.9

Lot 5: There were 40 units in this lot, at the end of the lot 110 units had been made at an average cost of 1130.6. Cost of this lot is:

Cumulative total cost for 110 units = 110(1130.6) =	124366
Minus cumulative total cost for 70 units	90048
Cost of this lot	34318
Average cost for this lot	858.0

It should be obvious the average cost or cost per unit for all units after the first lot is approximately 858 labor hours.

The correct analysis of this data is shown in Table 14-5.

Table 14-5

Lot No.	Lot Size	Lot Value	Non-Recurring	Recurring Cost	Average Recurring Cost for Lot
1	10	38578	30000	8578	857.8
2	15	12867		12867	857.8
3	25	21445		21445	857.8
4	20	17158		17158	857.9
5	40	34318		34318	858.0

Total cost of 110 units as recorded:

Nonrecurring Cost	30000
Recurring Cost	94366
Total Cost	124366

If one had to predict a follow-on lot the best prediction would be 858 hours per unit.

Note carefully: If one used only the cumulative average curve one would erroneously decide learning curves did apply.

Another case that could be misinterpreted and learning curve analysis erroneously applied is the case where costs are essentially fixed by production changes during periods of time. Once again preparation of a worksheet such as Figure 14-4 will detect that lot costs are the same, independent of lot size.

The proper preparation of worksheets will help greatly in determining whether or not the learning curve does apply.

Example of Use of Unit Curve:

The major use of learning curve analysis is to estimate costs. Assume we have to estimate the cost of 100 units of an airframe and we find the following:

- Manufacturing labor will have an 85% slope type unit curve with a first unit cost of 230,000 hours.

- Direct materials will have a 95% slope, a unit type curve, first unit cost $875.000 197X dollars.

- Engineering labor will have a 75% slope, a unit curve, unit one cost 964,000 hours.

- Tooling will have a 70% unit type curve, unit one cost, 1,230,000 hours.

- Quantity control labor will be 13% of manufacturing labor hours.

This is estimated or known:

EXAMPLE 14-1

Direct Cost	Slope of Unit Curve	Cost of Unit One (A)
Manufacturing Labor	85%	230,000 hrs
Materials	95%	875,000 197X dollars
Engineering Labor	75%	964,000 hrs
Tooling	70%	1,230,000 hrs

In each case the lot plot point will be 33.3. Since we have one lot of 100 units the lot midpoint will be 100/3, and with a first lot the lot midpoint and plot point are the same. For the lot:

$$\text{Total Cost} = 100(A)(33.3)^b$$

$$\text{and } b = \log \text{"slope"}/\log 2$$

One way to estimate the costs would be to use a calculator. In this case an HP-65.

To estimate manufacturing cost:

On Keyboard Enter	Press	Read
Nothing	DSP.8	0.00000000
33.3	ENTER	33.3
.85	Log	-.07058107
2	Log	.30103000
Nothing	$\div$	-0.23446525
Nothing	y^x	.43958182
230,000	X	101103.8195
100	X	10110382

hence the estimate of direct labor hours in 100 units this airframe is 10,110,382 at $10.00 an hour (197X rate) direct labor cost would be $101.102820 million or 101.1 million 197X dollars.

For direct material costs we will need:

$$100(875000)(33.3^b)$$

and $b = \log .95/\log 2 = -0.0740058$

the estimate (made with the same HP65 calculator) is:

$$100(875000)(33.3^{-.07400058})$$
$$= 100(875000)(.77150411) = 67.5 \text{ million 197X dollars}$$

Engineering labor we need:

$$100(964000)(33.3^b)$$

and $b = \log .75/\log 2 = -.4150375$

the estimated cost is:

22501195 engineering hours.

The tooling hours are:

$$100(1230000)(33.3^b)$$

and $b = \log .7/\log 2 = -.51457317$

the estimated cost is:

20253317 hours

The final estimate is quality control labor = 13% mfg labor = 13% or 10,110,382 or 1,314,350 hours.

Another way to make the estimates is to use a computer program. In this case the LEARN* program on the General Electric system was used. Results are:

GE Estimate	HP 65 Estimate Above	$% Difference $\frac{GE}{HP65}$ *100
Labor (hrs) 10,063,390	10,110,382	+0.46
Material ($197X) 67,013,090	67,506,610	+0.74
Engineering (hrs) 23,308.156	22,501,195	-3.46
Tooling (hrs) 21,883,572	20,253,317	-7.44

Certainly, the entries made with the HP65 calculators are workable estimates.

The third and simplest way to make the estimate is to use log-log paper. For example, to estimate labor hours you should calculate the values on the 85% learning curve.

(Carry as many digits as feasible)

Unit	Value	Source
1	230,000	given
2	195,000	85% of unit 1 value
4	166,200	85% of unit 2 value
8	141,300	85% of unit 4 value
16	120,100	85% of unit 8 value
32	102,100	85% of unit 16 value
64	86,700	85% of unit 32 value

Then plot and connect them as shown in Figure 14-8. Read the value at X = 33.3. It is (very close to) 101,000. This is the average value for the lot so 100 units time the average value--101,000 times 100 = 10,100,000 which value is close to our other estimates and was very easy to get, a piece of graph paper, plus straight edge and pencil were all the tools needed.

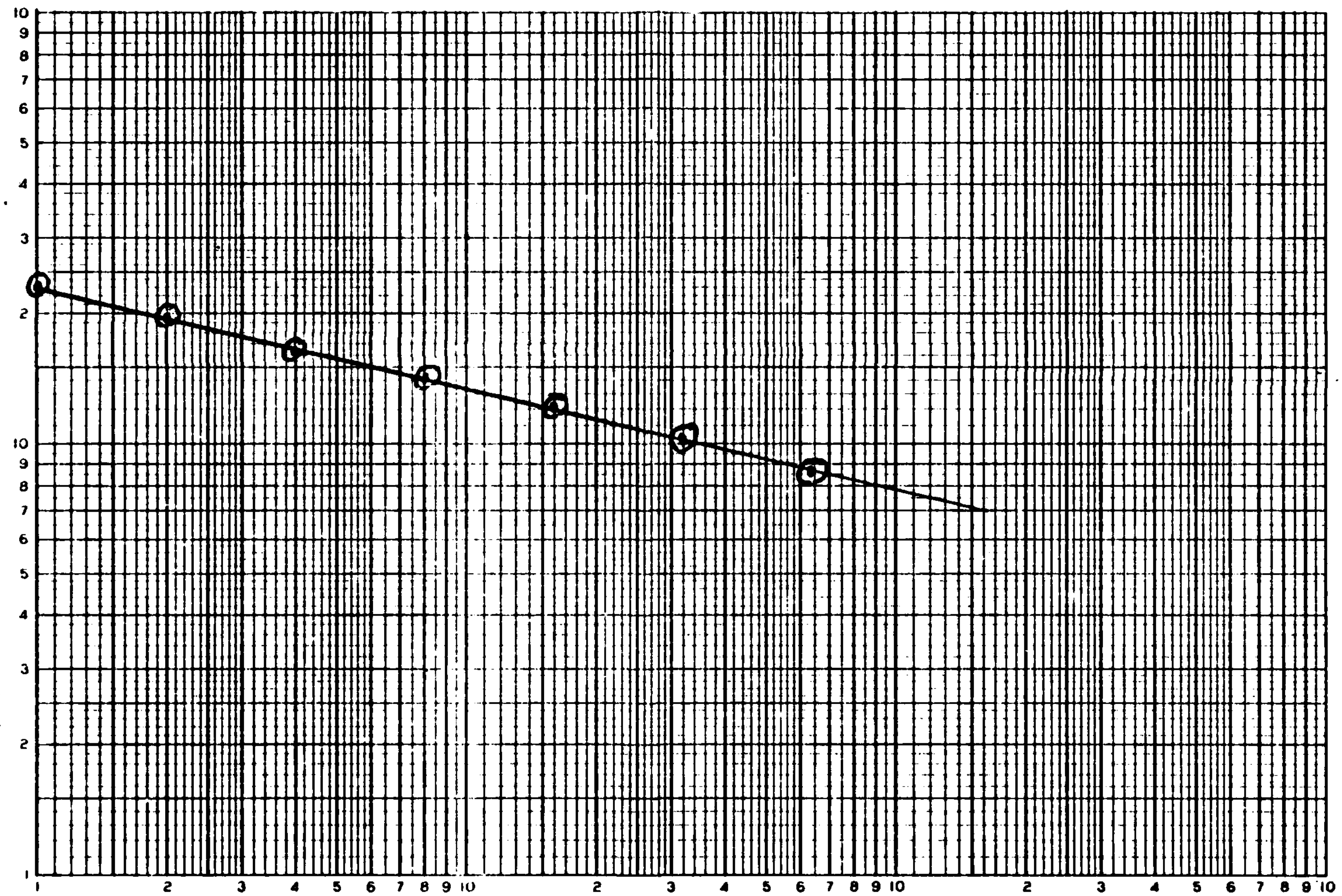
<u>To Estimate Costs of a Follow-on Lot (UC)</u>:

To estimate the costs of a lot other than the first, the same type of work is done.

As an example: Given the following data estimate the cost of a lot of 25 units; units 76 to 100 inclusive.

Figure 14-8

Plot of Unit Curve for Manufacturing
Labor from Example 14-1



EXAMPLE 14-2:

Direct Cost	Slope of Unit Curve	Cost of Unit One (A)
Manufacturing Labor	80%	250,000 hrs
Materials	93%	850,000 $198X
Engineering Labor	77%	950,000 hrs
Tooling	73%	120,000 hrs

The lot contains 25 units, the midpoint is 25 ÷ 2 = 12.5. Previous production is (or will be) 75 units so the plot point is 75 + 12.5 or 87.5.

Each estimate will be from an equation.

Cost of Lot = (Number of units in lot) (Cost of unit one) (Plot point raised to b)

$$TC_2 = 25(A)(87.5^b)$$

The appropriate b's are

for 80%: $\log(.80) \div \log(2) = -.0321928$
90%: $\log(.93) \div \log(2) = -0.104697$
77%: $\log(.77) \div \log(2) = -0.377070$
73%: $\log(.73) \div \log(2) = -0.454032$

The estimated costs of lot are calculated:

Manufacturing labor

$$\begin{aligned} \text{Cost} &= (25)(250{,}000)(87.5^{-0.321928}) \\ &= (25)(250{,}000)(.626149) \\ &= 1{,}481{,}500 \text{ hours approximately} \end{aligned}$$

$$\begin{aligned} \text{Materials} &= (25)(950{,}000)(87.5^{-0.377070}) \\ &= (25)(950{,}000)(.185237) \\ &= 4{,}399{,}400 \text{ hours approximately} \end{aligned}$$

$$\begin{aligned} \text{Tooling hours} &= (25)(120{,}000)(87.5^{-0.454032}) \\ &= (25)(120{,}000)(.131300) \\ &= 393{,}900 \text{ hours approximately} \end{aligned}$$

These estimates for lots after the first lot are more nearly those of high sophistication computer programs. Compare the outputs:

Estimate	GE Computer Estimate (Rounded to same no. of digits as desk calculator)	Desk Calculator Estimate	PCT Diff
Mfg labor	1,480,900	1,481,500	0.04%
Materials	13,302,900	13,305,700	0.02%
Engineering labor	4,397,700	4,399,400	0.04%
Tooling hours	393,800	393,900	0.03%

Figure 14-9 shows the graphs of the manufacturing labor and material curves. Unit 87.5 represents the average cost for the lot, the cost of the lot is the average multiplied by the number of units in the lot.

For manufacturing labor I read:

$$\overline{Y}_L = 59{,}200 \text{ times } 25 = 1{,}480{,}000$$

For direct material:

$$\overline{Y}_L = 528{,}000 \text{ times } 25 = 13{,}200{,}000$$

Figure 14-10 shows the engineering labor and tooling hours curves from which:

For engineering labor

$$\overline{Y}_L = 175{,}000 \text{ hrs times } 25 \text{ units} = 4{,}386{,}000$$

For tooling hours

$$\overline{Y}_L = 15{,}700 \text{ times } 25 = 393{,}000$$

A difficulty with graphs is that you cannot read them to more than three digits so these estimates look worse than those calculated. However, the precision from the calculation is more apparent than real--it's a function of the calculator word length. There are only three digits in the cost of unit one so not more than three digits are significant.

<u>Cumulative Average (CA) Formulation</u>:

If the cumulative average curve is the form of the learning curve to be used, then the estimation is different. With the unit curve we used a number near the middle of the lot because the average was in the middle of the lot.

When the cumulative average formulation describes the cost pattern the average for X units is plotted at X. With an 80% curve one has:

$$\overline{Y}_X = AX^{-.322928}$$

If unit one cost 1500 hours then:

$$\overline{Y}_5 = 1500(5^{-.321928}) = 1500\ (0.595637)$$
$$\approx 893$$

893 is a good estimate of the average cost. If further calculations are to be attempted we should carry more digits, say use 893.456 as the average and then the lot will cost 5(893.456) = 4467 or better 4470.

Figure 14-9

Plot of Unit Learning Curve for Manufacturing Labor and Direct Materials for Example 14-2

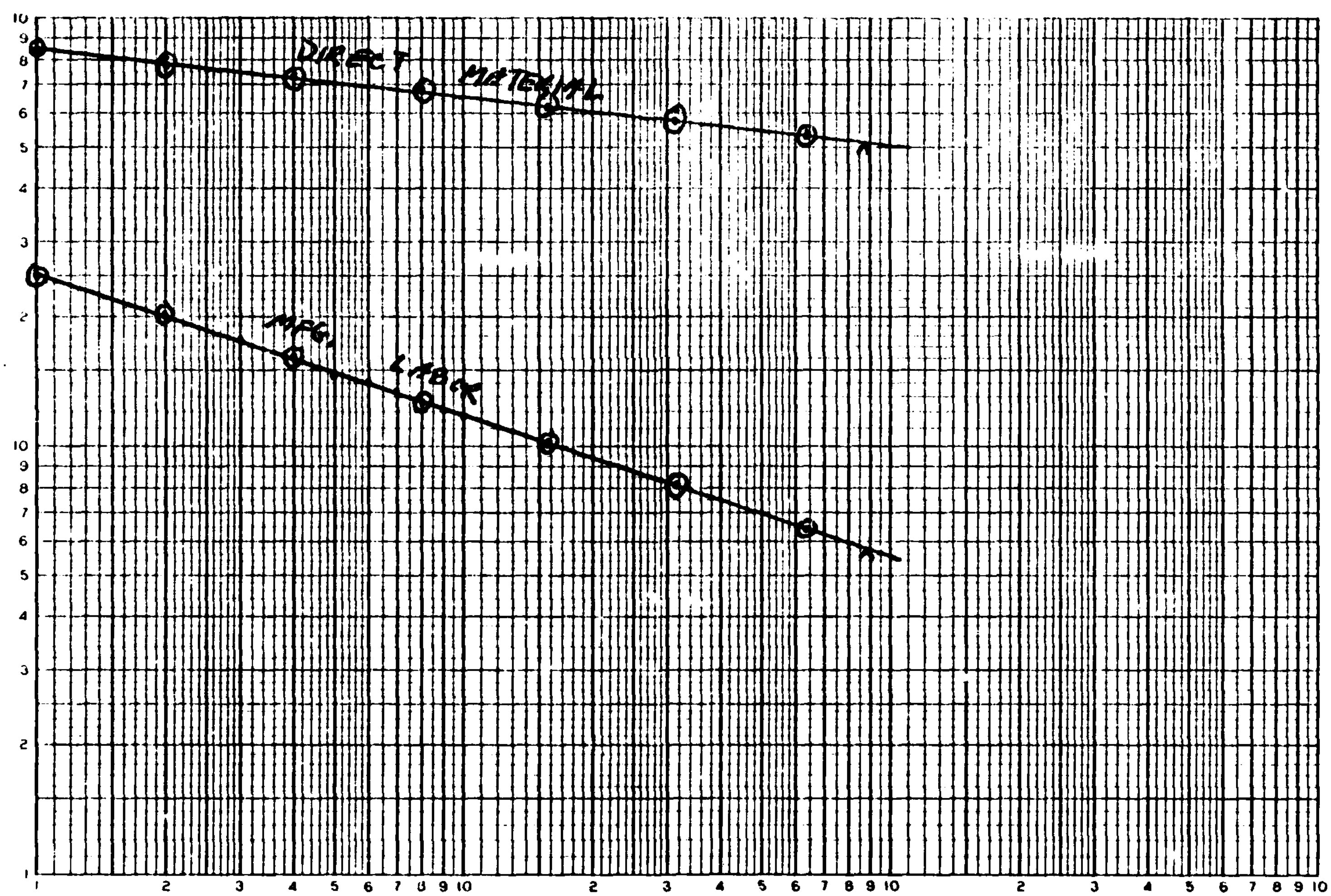


Figure 14-10

Plot of Unit Learning Curve for
Engineering Labor and Tooling
Hours for Example 14-2

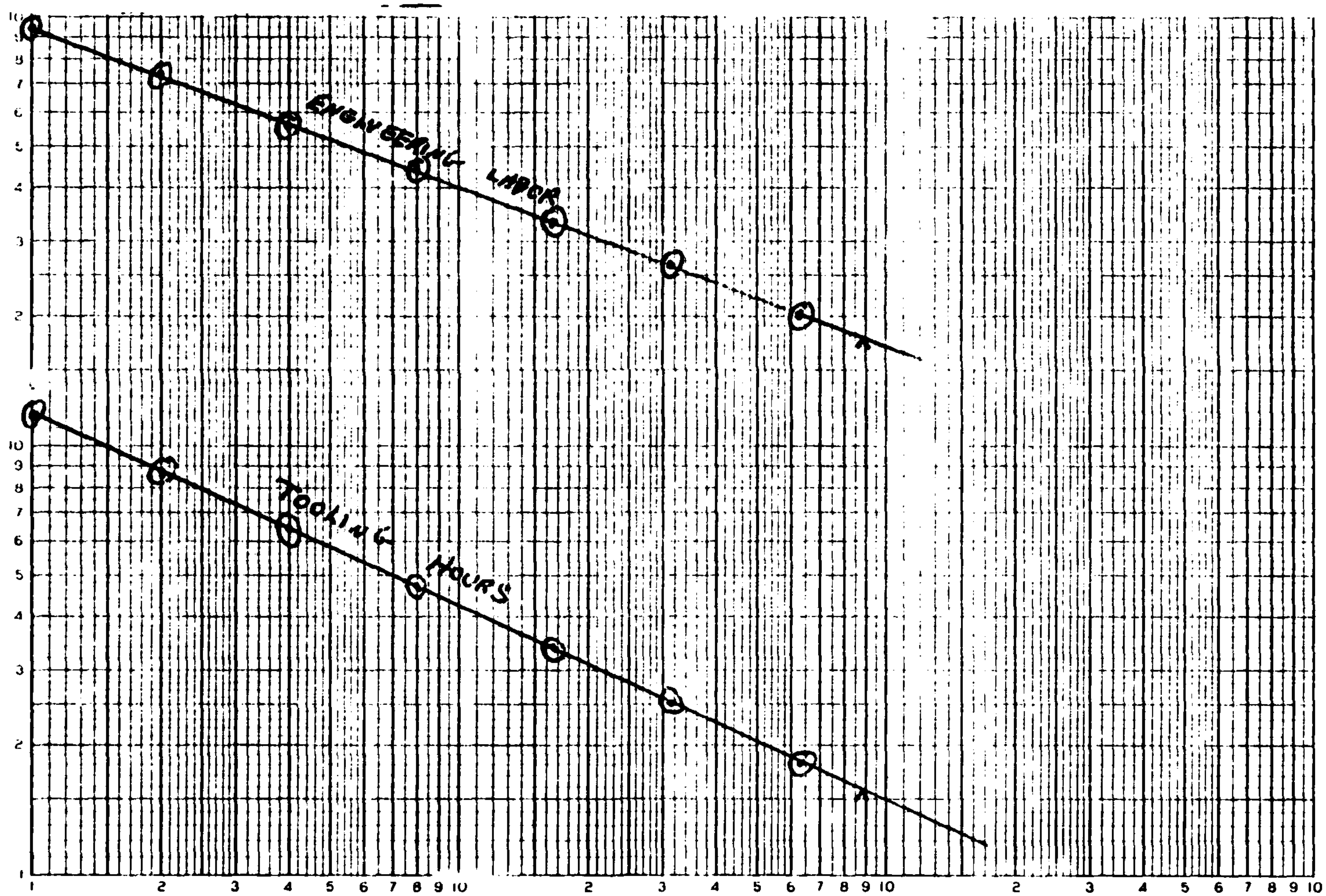


Table 14-6 describes an 80% cumulative average curve with a unit one cost of 2500 hours.

Table 14-6

Unit (X)	Average Cost of X units $\bar{Y}_X = 2500X^b$	Total Cost of X units $TC_X = X2500X^b$	Cost of X Unit $TC_X - TC_{(X-1)}$
1	2500	2500	2500
2	2000	4000	1500
3	1755	5266	1266
4	1600	6400	1134
5	1489	7446	1045
6	1404	8425	980
7	1336	9354	928
8	1280	10240	886
9	1232	11091	851
10	1191	11913	821

As with the unit curve, there are three ways to make estimates of the costs of a lot given the CA formulation--one may use a graph, an electronic calculator, or a computer program.

EXAMPLE 14-3:

A very new ultra modern fighter airframe has the following cost pattern.

Learning Curve

Cost	Form	Slope	First Unit Cost
Mfg Labor	Cum Av	82%	1,300 thousand hrs
Dir. Material	Cum Av	93%	5,215 thousand 198X dollars
Engineering	Cum Av	77%	228 thousand hrs
Tooling	Cum Av	75%	325 thousand hrs

Estimate cost of a first lot of 25 airframes.

The relevant exponents (b's) are:

82%:b = log .82/log 2 = -0.286304
93%:b = log .93/log 2 = -0.104697
77%:b = log .77/log 2 = -0.377070
75%:b = log .75/log 2 = -0.415038

Using an electronic calculator one estimates the cost of the 25 airframes as:

$$\text{Direct Labor} = (1300)(25^{-0.286304})(25) \text{ thousand hours}$$
$$= 12{,}931 \text{ thousand hours}$$

$$\text{Direct Material} = (5215)(25^{-0.104697})(25) \text{ thousand 198X dollars}$$
$$= 93{,}075 \text{ thousand 198X dollars}$$

$$\text{Engineering} = (228)(25^{00.377070})(25) \text{ thousand hours}$$
$$= 1693 \text{ thousand hours}$$

$$\text{Tooling} = (325)(25^{-0.415038})(25) \text{ thousand hours}$$
$$= 2136 \text{ thousand hours}$$

To graph these curves one uses exactly the same technique as with the unit curve but estimates of cost are made differently.

To graph the Manufacturing Labor Curve get some points on the curve; how many is a matter of choice but X as large as 64 or 128 (depending on the size of the paper) is desirable.

Cumulative Average Curve
for Manufacturing Labor From Example Three

Unit	Cumulative Average Cost (thousands of hours)	Source
1	1300	Given
2	1066	82% of value for unit 1
4	874	82% of value for unit 2
8	717	82% of value for unit 4
16	588	82% of value for unit 8
32	482	82% of value for unit 16
64	395	82% of value for unit 32

Figure 14-11 shows the learning curve for manufacturing labor and direct materials. Figure 14-12 is the plot of the learning curve for engineering labor and tooling hours. Using the plots to estimate the costs for this lot of 25 units we have:

Manufacturing labor = (518)(25) = 12,950,000 hours
Direct materials = (3625)(25) = 90,625,000 dollars
Engineering labor = (67.5)(25) = 1,687,500 hours
Tooling = (85.8)(25) = 2,145,000 hours

As in other cases the results from graphing and calculator operations are very close.

Figure 14-11

Cumulative Average Learning
Curves for Manufacturing Labor
and Direct Materials From
Example 14-3

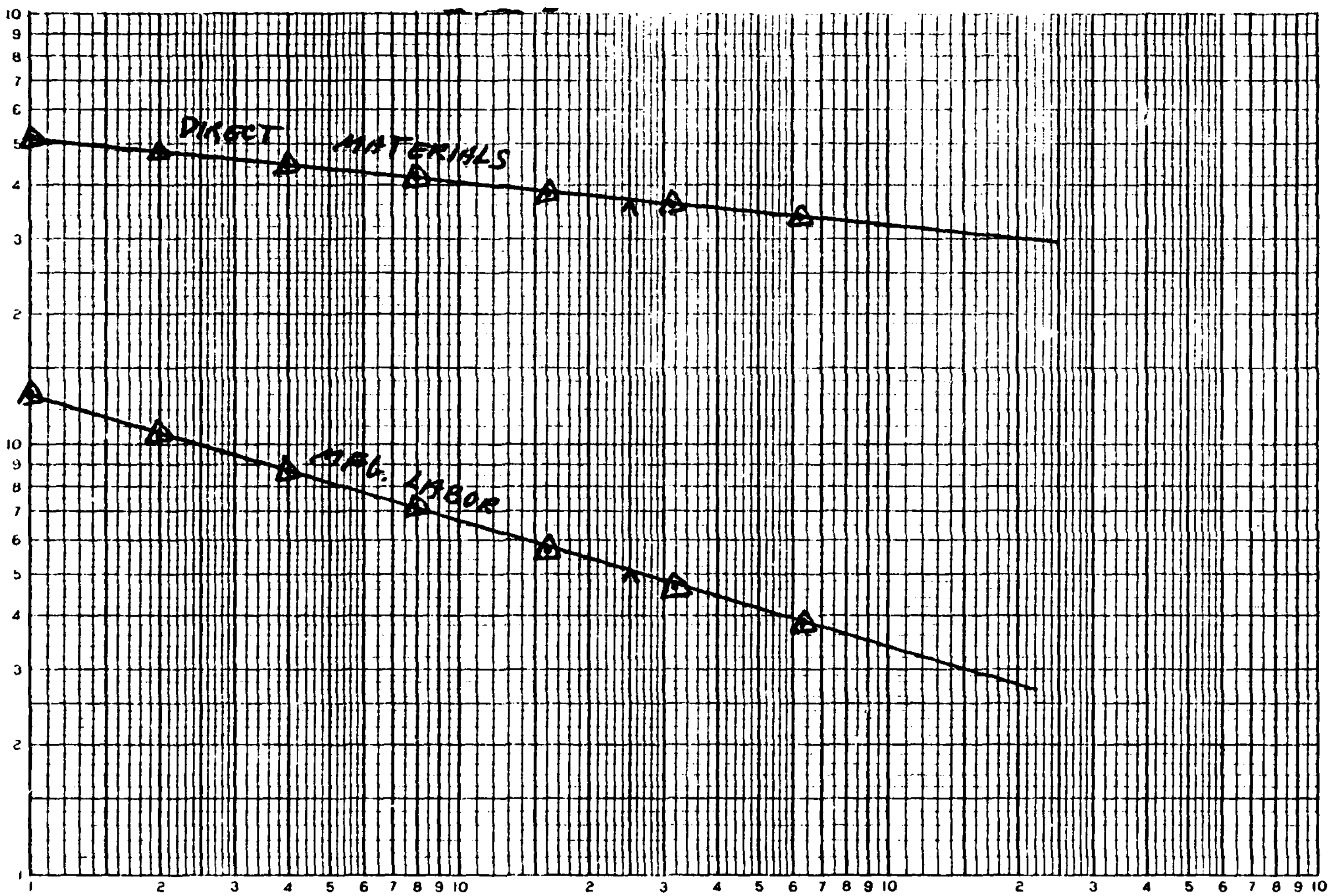
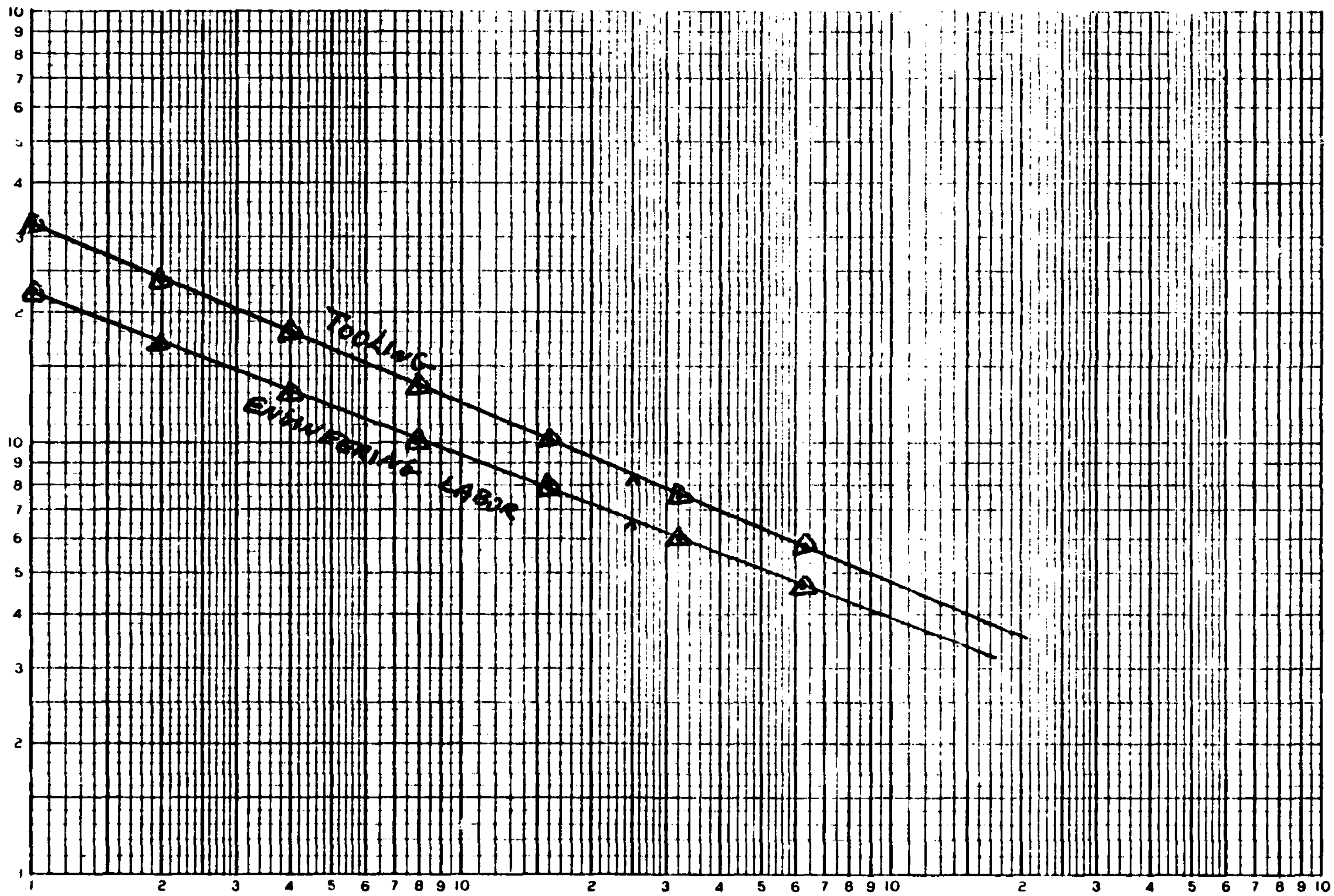


Figure 14-12

Cumulative Average Learning
Curves For Engineering Labor
And Tooling From Example 14-3



If you wish to estimate the cost of a lot other than the first lot, you must remember that cumulative average times the number of units gives total cost so you must subtract the cost of the previous lots.

For example, if we wish to estimate the cost of a lot of 25 units, (say units 76-100) using an electronic calculator:

$$TC_{76-100} = A\left[(100)(100^{b}) - (75)(75^{b})\right]$$

$$\text{Direct labor} = 1300\left[(100)(100^{-0.286304}) - (75)(75^{-0.286304})\right]$$
$$= 1300(26.754202 - 21.788319)$$
$$= 6{,}456{,}000 \text{ hours}$$

$$\text{Direct Material} = 5215\left[(100)(100^{-0.104697}) - (75)(75^{-0.104697})\right]$$
$$= 5215(61.745598 - 47.725224)$$
$$= 73{,}116{,}000 \text{ 198X dollars}$$

$$\text{Engineering Labor} = 228\left[(100)(100^{-0.377070}) - (75)(75^{-0.377070})\right]$$
$$= 228(17.614081 - 14.724207)$$
$$= 659{,}000 \text{ hours}$$

In each case we calculate the cost of 100 units and subtract the cost of 75 units to get the cost of the lot.

Using graphs we read the value at unit 100 which gives the cumulative average at unit 100, multiply by 100 to get the total cost of 100 units and subtract the reading at unit 75 times 75.

Using Figure 14-11 and 14-12 we get:

$$\text{Manufacturing Labor} = \left[(345)\ (100) - (376)(75)\right]$$
$$= 6{,}300{,}000 \text{ hours}$$

$$\text{Direct Material} = \left[(3190)\ (100) - (3300)(75)\right]$$
$$= 71{,}500{,}000 \text{ 198X dollars}$$

$$\text{Engineering Labor} = \left[(39.8)\ (100) - (44.1)(75)\right]$$
$$= 672{,}000 \text{ hours}$$

$$\text{Tooling} = \left[(47.3)\ (100) - (53.7)(75)\right]$$
$$= 702{,}000 \text{ hours}$$

Graphs give slightly different answers but as an estimate they should be very close to those answers calculated. Graphs are the simplest to use and they cost less than a sophisticated electronic calculator.

Learning curves can be useful tools for estimating direct costs. In spite of all the technological advances that have occured, a piece of graph paper remains a powerful and relatively easy tool to use in making these estimates.

Alternatives to graph paper, the modern electronic calculator and the computer, are nice but unnecessary and paper is much cheaper. Calculators and computers provide a lot of digits but the analyst must remember estimates are no better than the data on which they are based and a few digits are usually all we can trust.

Anybody can get a few packages of log-log paper and happily join the estimating community. Why don't you?

CHAPTER 15

ANALYSIS OF VARIANCE

INTRODUCTION

In Chapter 8 hypothesis tests were made on the mean of one population, and on the difference of means of two populations. This chapter discusses a technique for comparing r means and is called analysis of variance (ANOVA). The hypotheses to be tested are:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_r$$

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \ldots \neq \mu_r$$

If the null hypothesis is rejected, the conclusion is that at least one pair of means are significantly different, i.e., they come from different populations.

Recall that for testing one and two means, the t test was used. The t test could be used to test r means but the process would become cumbersome because each possible pair of means would have to be tested separately. For example, an experiment or process calling for the comparison of five means would require ten separate t tests (you haven't forgotten that 5 objects taken 2 at a time is 10).

ONE-FACTOR ANALYSIS OF VARIANCE

In testing hypotheses on the difference between means

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

the null hypothesis is rejected when the critical difference between sample means $(\bar{X}_1 - \bar{X}_2)$ is greater than

$$\pm \left[t_{\frac{\alpha}{2},\ d.f. = n_1 + n_2 - 2} \right] \cdot s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

and σ_1^2 and σ_2^2 are unknown but equal (see Case 3, page 8-21. In order to perform the foregoing test an estimate of the population variance was needed. This was found as

$$\hat{\sigma}^2 = s_p^2 \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $\hat{\sigma}^2$ is the estimator for σ^2 which is the variance of the two populations in question (assumed to be equal), and

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$$

In the test of hypotheses on r means, the F distribution is used, which is

$$F = \frac{ns_{\bar{x}}^2}{s_p^2}$$

Both numerator and denominator are estimates of the population variance σ^2. In this case, however, the denominator is referred to as a pooled variance and is not the same s_p^2 used in Case 3 of the t test. The pooled variance can be found by:

$$s_p^2 = \left[\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}\right] \cdot \left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

It can be readily be seen that $ns_{\bar{x}}^2$ is also an estimator of the population variance. Since s^2 is an estimator of σ^2 and $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$, then $\sigma^2 = \sigma_{\bar{x}}^2 \cdot n$ which is estimated by $n \cdot s_{\bar{x}}^2$. As will be demonstrated later, the actual computations for the test statistic $F = \frac{ns_{\bar{x}}^2}{s_p^2}$ is made easier (and more meaningful) through the use of sums of squares.

Single Factor ANOVA Model. A factor is an independent variable which has been isolated to determine its effect on a specified dependent variable. For example, a maintenance officer might be interested in determining the effect of maintenance technician skill levels (factor) on the amount of time needed to complete a job. In this case, the single factor, skill level, would consist of several factor levels, e.g., 3, 5, 7, etc. The dependent variable, time to complete a job, may be "dependent" upon the skill level of the technician doing the work.

The linear equation for the single factor ANOVA model is given by:

$$X_{ij} = \mu + R_i + e_{ij}$$

where X_{ij} = the value of the dependent variable being measured, e.g., the amount of time to complete a specific job by technician j with skill level i.

μ = overall mean; the average of all possible values of the dependent variable

R_i = row (or treatment) effect; in this example it is the effect of skill level on average time to complete the job

where $R_i = \mu_i - \mu$

and $\Sigma R_i = 0$ since under H_o,

$$\mu_1 = \mu_2 = \mu_3 \cdot \cdot \cdot = \mu_r$$

e_{ij} = random effect, the difference in X_{ij} values not accounted for by difference in skill level

where $e_{ij} = X_{ij} - \mu_i$

Assumptions for the ANOVA model are made on this random component; for statistical tests using this model to be valid, the e_{ij} must be:

1. Independent

2. Normally distributed for each treatment (row) subpopulation.

3. $E(e_{ij}) = 0$

4. Var $(e_{ij}) = \sigma^2$ (homogeneity)

ANOVA Table. In our example, the maintenance officer recognizes that the amount of time to complete a specific job fluctuates each time it is performed, and wishes to isolate the reasons for the differences in completion times. The ANOVA table below is a convenient way to show the sources of fluctuation (called variation in ANOVA).

Table 15-1

ANOVA Table

Source	Sum of Squares (SS)	Degrees of freedom (df)	Mean sum of Squares (MSS)	F
Explained (Between rows)	SS_r	r-1	$ns_{\bar{x}}^2 = \frac{SS_r}{r-1}$ (MSS_r)	$\frac{MSS_r}{MSS_e}$
Unexplained (Within rows or error)	SS_e	r(n-1)	$s_p^2 = \frac{SS_e}{r(n-1)}$ (MSS_e)	
Total	SS_t	nr-1		

The sources of variation are best illustrated by the equation using sums of squares:

$$\underbrace{\sum_i \sum_j (X_{ij} - \bar{\bar{X}})^2}_{\substack{\text{total}\\ \text{variation}\\ (SS_t)}} = \underbrace{n \sum_i (\bar{X}_i - \bar{\bar{X}})^2}_{\substack{\text{explained}\\ \text{variation}\\ (SS_r)}} + \underbrace{\sum_i \sum_j (\bar{X}_i - \bar{\bar{X}})^2}_{\substack{\text{unexplained}\\ \text{variation}\\ (SS_e)}}$$

where,

$\bar{\bar{X}}$ = overall sample mean found by

$$= \sum_i \sum_j \frac{X_{ij}}{\sum_i n_i} = \frac{\sum_i n_i \bar{X}_i}{\sum_i n_i}$$

$$\bar{X}_i = \sum_i \frac{X_i}{n_i}$$

The <u>total variation</u> represents differences between each observed value and the mean of all values. This total variation has two components. The <u>explained variation</u> component represents differences between the mean of each factor level (row) and the overall mean. This component or source of variation isolates the portion of the total variation which is explained or accounted for by the factor levels, i.e., how much of the total variation in completion time is explained by the fact that technicians of different skill levels accomplish the job.

The remaining source of variation, unexplained, is the portion of total variation which has not been explained by the differences of factor levels. This variation is said to be due to chance (random) or the result of other factors not identified in the model.

After the sources of variation have been isolated, an F-value is obtained by:

$$F = \frac{\text{explained variance}}{\text{unexplained variance}}$$

$$= \frac{\dfrac{SS_r}{r-1}}{\dfrac{SS_e}{r(n-1)}}$$

$$= \frac{MSS_r}{MSS_e}$$

A large value of F means that the explained variance is much larger than the random variance and that the hypothesis of no difference, i.e.,

$$H_o: = \mu_1 = \mu_2 = \mu_3 = \cdot \cdot \cdot = \mu_r$$

should be rejected.

<u>EXAMPLE 15-1</u>. Three road bedding machines are being tested by the Base Civil Engineer who is interested in knowing whether the average output differs among the machines. Output is measured as linear feet of roadbed produced per hour. The hypotheses to be tested would be:

$$H_o: = \mu_1 = \mu_2 = \mu_3$$

$$H_1: = \mu_1 \neq \mu_2 \neq \mu_3$$

The test will be conducted at a significance level of $\alpha = 0.05$. A sample of five hours' output is obtained for each of the machines. Table 15-2 shows the actual data in linear feet per hour.

Table 15-2

MACHINE HOURS

Machine \ Hours	j = 1	2	3	4	5
i = 1	54.6	45.7	56.7	37.7	48.3
2	53.4	57.5	54.3	52.3	64.5
3	56.7	44.7	50.6	56.5	49.5

The components of variation (in output) are computed below and placed in the ANOVA table form for analysis.

$$SS_r = \frac{\sum_{i=1}^{r} X_{i.}^2}{n} - \frac{X_{..}^2}{nr}$$

where $\sum_{i=1}^{r} X_{i.}^2$ = [total of row i]^2 = $\sum_{i=1}^{r} \left(\sum_{j=1}^{n} X_{ij}\right)^2$

$X_{..}^2$ = [total of all observations]^2 = $\left(\sum_{i=1}^{r} \sum_{j=1}^{n} X_{ij}\right)^2$

n = number of observations in each row (assumes row sample sizes are equal)

r = number of rows

$$SS_r = \frac{205,137}{5} - \frac{613,089}{5 \cdot 3} = 158.4$$

$$SS_t = \sum_{i=1}^{r} \sum_{j=1}^{n} X_{ij}^2 - \frac{X_{..}^2}{nr}$$

$$= 41,456.4 - \frac{613,089}{5 \cdot 3} = 583.8$$

$$SS_e = SS_t - SS_r \quad \text{or}$$

$$SS_e = \sum_{i=1}^{r} \sum_{j=1}^{n} X_{ij}^2 - \frac{\sum_{i=1}^{r} X_{i.}^2}{n}$$

$$= 41{,}456.4 - 41{,}027.4 = 429$$

Table 15-3

ANOVA TABLE

Source	SS	df	MSS	F
Explained (Between rows)	154.8	2	77.4	2.165
Unexplained (Within rows)	429	12	35.75	
Total	583.8	14		

To test the hypothesis that the row means are equal, a critical value of F (F_C) is obtained from a table of critical values of the F statistic for

$$F^{r-1}_{r(n-1),\ \alpha} \quad \text{or} \quad F^{2}_{12,\ \alpha = .05} \quad \text{which is 4.10}$$

Since $F \not> F_c$, H_O cannot be rejected and it must be concluded that the average outputs of the three machines are approximately equal.

Unequal Sample Sizes. The ANOVA Table can be constructed for cases in which the sample sizes are not equal as follows:

Table 15-4

Source	SS	df	MSS	F
Explained (Between rows)	$\sum_{i=1}^{r} \frac{\left(\sum_{j=1}^{n_i} x_{ij}\right)^2}{n_i} - \frac{\left(\sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij}\right)^2}{\sum_{i=1}^{r} n_i}$	r-1	$\frac{SS_r}{r-1}$	$\frac{MSS_r}{MSS_e}$
Unexplained (Within rows)	$SS_t - SS_r$	$\sum_{i=1}^{r} (n_i - 1)$	$\frac{SS_e}{\sum_{i=1}^{r} (n_i - 1)}$	
Total	$\sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij}^2 - \frac{\left(\sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij}\right)^2}{\sum_{i=1}^{r} n_i}$	$\sum_{i=1}^{r} n_i - 1$		

Multiple Comparisons. The overall F-test for the model is a test to determine whether there is at least one pair of means which are significantly different. To isolate which of the means are statistically different , a method of contrasting should be used. The method of H. Scheffe presented below is the most conservative of the a posteriori approaches to comparing means. Other methods such as Duncan's multiple range test, Tukey, and least significant difference (LSD) can be found in statistical applications texts.

The Scheffe method of contrasting is useful particularly when the researcher has not decided a priori the exact nature of contrasting to be done prior to obtaining the experimental data.

Pairwise comparisons are found by:

$$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm \left[\sqrt{F^{(r-1)}_{\alpha,n(r-1)}} \cdot \sqrt{MSS_e} \cdot \sqrt{\sum C_i^2 \cdot \frac{(r-1)}{n}}\right]$$

where

$\bar{x}_1 - \bar{x}_2$ = difference in means of samples (rows) 1 and 2 being tested

$F^{(r-1)}_{\alpha,n(r-1)}$ = tabled F value for confidence level of $1 - \alpha$ and $r - 1$ degrees of freedom in the numerator, and $n(r-1)$ degrees of freedom in the denominator.

MSS_e = mean sum of squares (error) from the ANOVA table

C = contrasting coefficient. In the equation above, $C_1 = 1$ and $C_2 = -1$. $\sum_i C_i \equiv 0$.

In the example above, the average output of machines 1 and 2 using $\alpha = .05$ would be:

$$\mu_1 - \mu_2 = (48.6 - 56.4) \pm \left[\sqrt{3.89} \cdot \sqrt{35.75} \cdot \sqrt{2 \cdot \frac{2}{5}}\right]$$

$$= -7.8 \pm 10.55$$

The 95 percent confidence interval thus constructed is:

$$-18.35 \leq \mu_1 - \mu_2 \leq 2.75$$

The relevant hypotheses which are tested in this example are:

H_o: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

Since the confidence interval on $\mu_1 - \mu_2$ contains the value zero, the null hypothesis cannot be rejected, i.e., there is no significant difference in the average outputs of roadbed machines 1 and 2. This is to be expected since the overall F-test resulted in nonrejection of H_o: $\mu_1 = \mu_2 = \mu_3$. It should be noted that by using Scheffe's technique for comparison, any number of means may be compared simultaneously without any loss in the level of significance.

TWO-FACTOR ANALYSIS OF VARIANCE

Two factor ANOVA enables the experimenter to identify, isolate, and test an additional source of variation. In the previous example, it was found that the differences in the outputs of the machines were not due to differences in the machines themselves, i.e., the machines factor was not a significant source of the variation in output. However, perhaps the machine operators, a second factor, may have an effect on machine output.

Extending the example, suppose there were five operators, each having worked the three machines. The data table is presented as before, this time considering the five operators.

Table 15-5

Machine \ Operator	j = 1	2	3	4	5
i = 1	54.6	45.7	56.7	37.7	48.3
2	53.4	57.5	54.3	52.3	64.5
3	56.7	44.7	50.6	56.5	49.5

The two factor ANOVA model is:

$$X_{ij} = \mu + R_i + C_j + e_{ij}$$

where

X_{ij} = value of the dependent variable being measured

μ = overall mean

R_i = row (treatment) effect

and

$R_i = \mu_i - \mu$

$\sum R_i = 0$

C_j = Column (block) effect

and

$C_i = \mu_j - \mu$

$\sum C_j = 0$

e_{ij} = random effect; the difference in X_{ij} not accounted for by differences in row and column observations.

The assumptions for the two-factor model are the same as those for the one-factor ANOVA and will not be repeated here. The two-way ANOVA Table is presented below:

Table 15-6

TWO-WAY ANOVA TABLE

Source	SS	df	MSS	F
Explained (Between rows, or treatment)	SS_r	r-1	$\frac{SS_r}{r-1} = MSS_r$	$\frac{MSS_r}{MSS_e}$
Explained (Between columns, or block)	SS_c	c-1	$\frac{SS_c}{c-1} = MSS_c$	$\frac{MSS_c}{MSS_e}$
Unexplained (error)	SS_e	(r-1)(c-1)	$\frac{SS_e}{(r-1)(c-1)} = MSS_e$	
Total	SS_t	rc-1		

Similar to the one-factor model, the sums of squares are obtained by:

$$\sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij}-\bar{X})^2 = c\sum_{i=1}^{r}(\bar{X}_{i.}-\bar{\bar{X}})^2 + r\sum_{j=1}^{c}(\bar{X}_{.j}-\bar{\bar{X}})^2 + \sum_{i=1}^{r}\sum_{j=1}^{c}(X_{ij}-\bar{X}_{i.}-\bar{X}_{.j}+\bar{\bar{X}})^2$$

$$\begin{matrix} \text{Total} & & \text{Row} & & \text{Column} & & \text{Unexplained} \\ \text{Variation} & = & \text{Variation} & + & \text{Variation} & + & \text{Variation} \\ SS_t & & SS_r & & SS_c & & SS_e \end{matrix}$$

SS_t and SS_r are found the same as in the one-factor model. SS_e is best calculated by:

$$SS_e = SS_t - SS_r - SS_c$$

where

$$SS_c = \frac{\sum_{j=1}^{c} X^2_{.j}}{r} - \frac{X^2_{..}}{rc}$$

and

$$X^2_{.j} = [\text{column totals}]^2 = \sum_{j=1}^{c}\left(\sum_{i=1}^{r} X_{ij}\right)^2$$

EXAMPLE 15-2. Continuing the illustration of the roadbedding equipment, the following ANOVA table for the two-way classification is constructed:

Table 15-7

ANOVA TABLE

Source	SS	df	MSS	F
Between rows	154.8	2	77.4	13.1
Between columns	381.7	4	95.4	16.2
Unexplained	47.3	8	5.9	
Total	583.8	14		

We can now test whether there is a discernible difference in machines and/or operators in the presence of each other. In other words, while testing for differences in machines, we are taking into account the differences in operators.

To test for differences in machines:

$$F = \frac{MSS_r}{MSS_e} = \frac{77.4}{5.9} = 13.1$$

The tabled value for $F^2_{8,\alpha=.05} = 4.46$ and the hypothesis of no difference in machines is rejected. Thus, in the presence of the operators, there is a significant difference in the average output of the machines.

To test for differences in operators:

$$F = \frac{MSS_c}{MSS_e} = \frac{95.4}{5.9} = 16.2$$

Obviously, the hypothesis of no difference in operators can also be rejected. Thus, there is a significant difference in operators in the presence of the machines.

Multiple Comparisons. In a manner similar to that shown in the one factor ANOVA model, the H. Scheffe method of contrasting can be done with column (block) effects as well as row (treatment) effects. The Scheffe confidence interval is given by:

$$\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \pm \left[\sqrt{F^{(r-1)}_{\alpha,(r-1)(c-1)}} \cdot \sqrt{MSS_e} \cdot \sqrt{\frac{r-1}{c}\sum_{j=1}^{r} C_j^2}\right]$$

difference in column means

The results are interpreted the same--if the conficence interval contains zero, there is no statistical difference between the pair of means.

"Collapsing" the ANOVA Table. After the F-tests on treatment and blocking effects have been performed and it is found that one of the factors is not significant, i.e., the null hypotheses cannot be rejected, the ANOVA Table should be reconstructed into a one-factor classification. For example, if the blocking effects are not significant, the explained variation (between columns) should be added to the residual or unexplained variation. The model should then be tested as a one-factor ANOVA for treatments (row effects) only. This is commonly referred to as collapsing the ANOVA Table and is illustrated below:

Table 15-8

TWO-FACTOR TABLE

Source	SS	df	MSS	F
Between rows (treatment)	260	4	65	14.8
Between columns (Block)	144	8	14.4	3.3
Unexplained (Residual or error)	44	10	4.4	
Total	448	22		

At $\alpha = .01$, block or column effects are not significant and should be added back into the residual. The following table results.

Table 15-9

ONE-FACTOR "COLLAPSED" TABLE

Source	SS	df	MSS	F
Between rows (treatment	260	4	65	6.2
Within rows (Error or residual)	188	18	10.44	
Total	448	22		

The hypothesis on row differences should then be retested to determine whether it remains significant. Note that in the collapsed table, the degrees of freedom as well as SS_C has been added into the residual effects. The treatment effects (SS_r) and total variation remain the same.

OTHER EXPERIMENTAL DESIGNS

Up to this point, we have been considering one and two factor fixed effects ANOVA models. Fixed effects models are those in which all of the levels of the factor(s) have been included in the experiment. In the example on the roadbedding equipment, we had to assume only three machines and only five operators were available. If more machines and operators were available and only three and five had been selected respectively, the example would have been a random effects model and the calculations for sources of variation would have been slightly different.

In the previous models, it was also assumed there was no interaction among the main factor effects. Such interaction would occur, for example, if some of the operators liked a particular machine and others disliked it. To examine the interaction effects, more than one observation per cell is required. Again in the previous example, each operator worked each machine one time and the output was observed and recorded. To analyze the interactions, each operator would have to run each machine several times. The process of obtaining more than one observation per cell is called replication. The advantage of replication is that more information can be obtained. However, this advantage must be weighed against the cost and accuracy of the data to be obtained.

Two-Factor ANOVA with Interaction. The equation for the model is given as:

$$X_{ijk} = \mu + R_i + C_j + (RC)_{ij} + e_{ijk}$$

where

X_{ijk} = observed value of the dependent variable

R_i = row effect

C_i = column effect

$(RC)_{ij}$ = effect due to the interaction of row and column factors

e_{ijk} = effect due to random fluctuation, experimental error.

R_i and C_I are called main effects since these are the main factors selected for isolation in the experiment. Hypotheses are tested for each of the main effects as well as the interaction effect to determine statistical significance. The assumptions for this model are made on the random error term, e_{ijk}, and are the same as presented earlier. The ANOVA

Table for the two-factor ANOVA with interactions is given below:

Table 15-10

Source	SS	df	MSS	F
Main effects				
Rows	$SS_r = n \cdot c \sum (\bar{X}_{i..} - \bar{\bar{X}})^2$	r-1	$\frac{SS_r}{r-1} = MSS_r$	$\frac{MSS_r}{MSS_e}$
Columns	$SS_c = n \cdot r \sum (\bar{X}_{.j.} - \bar{\bar{X}})^2$	c-1	$\frac{SS_c}{c-1} = MSS_c$	$\frac{MSS_c}{MSS_e}$
Interaction	$SS_{rc} = n\sum\sum (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{\bar{X}})^2$	(r-1)(c-1)	$\frac{SS_{rc}}{(r-1)(c-1)} = MSS_{rc}$	$\frac{MSS_{rc}}{MSS_e}$
Unexplained (Residual)	$SS_e = \sum\sum\sum (X_{ijk} - \bar{X}_{ij.})^2$	rc(n-1)	$\frac{SS_e}{rc(n-1)} = MSS_e$	
Total	$SS_t = \sum\sum\sum (X_{ijk} - \bar{\bar{X}})^2$	*nrc-1		

*n = number of observations per cell, i.e., replications.

EXAMPLE 15-3. A DOD repair facility is interested in studying the effects of several factors in a repair process which machines large castings on a hand-controlled lathe. The DOD wants to know if the type of cutting tool and angle setting of the tool significantly affect the amount of time it takes to machine a casting. Two types of cutting tools and two angle settings are used. For the experiment it has been decided that two castings will be lathed for each combination of cutting tool and angle setting, leaving 2·2·2 or 8 castings to be turned. The time to complete each casting will be recorded in minutes.

In order that the experiment remain unbiased, the order in which the castings were lathed was determined by a coin toss with the following results:

Table 15-11

	Tool Type			
	1 Angle		2 Angle	
	A	B	A	B
Order of Lathing	6	1	3	4
	8	2	7	5

The first casting, therefore, was using Tool Type 1, set at cutting angle B. The model to be used is:

$$X_{ijk} = \mu + T_i + A_j + (TA)_{ij} + e_{ijk}$$

where

X_{ijk} = cutting time in minutes

μ = true average cutting time for the process

T_i = the effect on cutting time of tool type i with $i = 1,2$

A_j = effect on cutting time of angle setting j with $j = A,B$

$(TA)_{ij}$ = interaction effect of tool type with cutting angle on average cutting time

e_{ijk} = random fluctuation; since there are two replications, $k = 1,2$

Data from the experiment are recorded in minutes in the following table.

Table 15-12

TOOL TYPE (T_i)

Replication	1			2		
	Angle (A_i)			Angle (A_i)		
	A	B	Sum	A	B	Sum
1	29	27	56	25	32	57
2	28	23	51	26	27	53
Sum (A_i)	57	50	107	51	59	110
Sum (T_i)	107			110		
Total	224					

Using the data from the table, the components of variation are completed from the following:

$$T = \sum_{ijk}\sum\sum X_{ijk} = \text{sum of all observations} = X...$$

n = number of replications

r = number of rows

c = number of columns

$$SS_t = \sum_{ijk}\sum\sum X^2_{ijk} - \frac{T^2}{nrc}$$

$$SS_r = \frac{\sum X^2_{i..}}{nc} - \frac{T^2}{nrc}$$

$$SS_c = \frac{\sum X^2_{i.j.}}{nr} - \frac{T^2}{nrc}$$

$$SS_e = \sum\sum\sum X^2_{ijk} - \frac{\sum_{ij}\sum X^2_{ij.}}{n}$$

Interaction sum of squares can be obtained as the rema'nder by

$$SS_{rc} = SS_t - SS_r - SS_c - SS_e$$

Table 15-13

ANOVA TABLE

Source	SS	df	MSS	F
Main Effects				
Between Tools	1.12	1	1.12	.2
Between Angles	.12	1	.12	.02
Interactions	28.13	1	28.13	5.2
Unexplained	21.5	4	5.375	
Total	50.87	7		

The critical value for $F^1_{4,\alpha = 0.5}$ is 7.71. Therefore, no significant effect on cutting time has been found due to tool type, cutting angle, or their interaction. If the test were mode at $\alpha = .01$, only the interaction would have been significant.

In summary, ANOVA can be a powerful tool in isolating the sources of variation for selected measureable variables. The processes presented in the previous pages are not all-inclusive of the wide range of ANOVA techniques. The reader should consult texts in experimental design for further reference.

CHAPTER 16

LINEAR PROGRAMMING

INTRODUCTION

Linear programming (LP) is a mathematical technique for determining the manner in which limited available resources can be allocated among competing alternatives in such a way as to optimize a particular objective while satisfying a number of constraints. As such, LP is one of a number of management science methods known collectively as "constrained optimization techniques." Because the problem of resource allocation is common to all organizations and, to varying degrees, an activity of concern to managers at all levels of responsibility, LP is one of the most widely used management science techniques.

Applications

The following list, which is by no means exhaustive nor even necessarily comprehensive, provides a representative inventory of prototype situations to which LP has been successfully applied:

1) Accounting
2) Agricultural Planning
3) Air Pollution Control
4) Airline Scheduling
5) Assignment and Scheduling
6) Capital Budgeting
7) Capital Expansion
8) Contract Award
9) Cargo Loading
10) Diet/Menu Planning
11) Environmental Protection
12) Equipment Acquisition
13) Financial Mix
14) Highway Planning
15) Hospital Administration
16) Inspection
17) Material Blending
18) Media Mix (Marketing)
19) Personnel/Manpower Planning
20) Plant Location
21) Police Patrol Sector Planning
22) Portfolio Selection
23) Product Mix
24) Production Scheduling
25) Production-Vendor Selection

26) Project Management
27) Public Works Management
28) Regional Planning
29) Solid Waste Management
30) Structural Analysis
31) Traffic Flow Analysis
32) Transportation-Distribution
33) Trim Loss
34) Water Resource Management
35) Welfare Allocation

Characteristics and Assumptions

Linear Programming can, in general, be applied to any decision situation having the following characteristics:

1. A limited quantity of economic resources (e.g., labor, capital, time, equipment, material, etc.) is available for allocation.

2. The resources are used in, or have utility for, the production of goods and/or services.

3. There are alternative courses of action, i.e., there are two or more ways in which the resources can be used. Each is called a solution or a program.

4. There is a clearly identifiable criterion by which alternative programs can be compared (e.g., cost, profit, etc.). Each product or service to which resources are allocated yields a return or makes some contribution in terms of the stated criterion.

5. The allocation is usually restrained by several limitations or requirements termed <u>constraints</u>.

6. It is possible to describe the decision situation as a mathematical model, i.e., the criterion and constraints can be expressed as (linear) mathematical equalities and/or inequalities.

In addition to meeting these characteristic conditions, decision situations for which LP is appropriate must satisfy the following basic assumptions:

1. <u>Certainty</u>. All data concerning the decision situation is assumed to be known with certainty, i.e., the model is deterministic.

2. <u>Proportionality</u>. All relationships (equalities and/or inequalities) in the model are linear, i.e., the criterion and constraints change proportionally to the level of each activity.

3. <u>Additivity</u>. The total utilization of each resource is determined by <u>adding</u> together that portion of the resource required for the production

of each of the various products or activities. Similarly, the net change in the criterion value is the sum of the individual contributions associated with each of the products or activities. The assumption of proportionality guarantees linearity if and only if joint effects or interactions are nonexistent.

4. Divisibility. The variables in the model are continuous and can assume fractional values. Divisibility means that a fractional value of the decision variables makes sense, and the activities are not restricted to whole numbers.

5. Nonnegativity. Negative activity levels (i.e., negative allocations are not permissible. All variables must assume nonnegative values.

6. Independence. Complete independence of coefficients is assumed, both among activities and among resources. For example, the cost of providing one unit of service A has no effect on providing one unit of service B.

MODEL FORMULATION

Mathematical Modeling

A model is, in general, a simplified abstraction or representation of some real system or phenomenon in which an individual is interested and about which that person wants to make a decision. A mathematical model is a system of mathematical equalities and/or inequalities that represent the decision situation of interest. As previously noted, an LP model is a mathematical model in which the component variables are continuous and nonnegative, the model parameters or constants are assumed to be known with certainty (i.e., they are deterministic), and all relationships linking the variables and/or constants are linear, i.e., they are proportional and additive.

Formulating an LP Model

In formulating an LP model to help improve decision-making effectiveness, it is helpful to use some type of conceptual framework to analyze the decision situation. One simple and convenient framework is to think about the problem as a system, i.e., as a set of interrelated and interacting components. Therefore, to describe any decision system, we need to identify:

1. the components in the system (i.e., the variables and constants) and

2. the relationships linking or existing among the components of the system.

System Components:

The first step in the analysis of any decision situation is so basic that it is often overlooked. Who is the manager or decision taker relative

to the problem at hand? In many cases, the analyst and manager/decision taker are the same. In other situations, however, the analyst acts in a staff role relative to the manager for whom the analysis is being performed. This determination is a vital prerequisite to constructing a decision system model that accurately and completely captures the values and range of influence or control of the manager.

In any decision situation, there is typically (at least) one variable that is the primary focus of the manager's attention. It is the thing that he or she wants to control. It is a system quality or characteristic in which the decision taker places some value. It is also a yardstick or criterion by which managers judge the performance or health of the real system with which they are concerned and responsible. In an LP decision system model, this component is commonly referred to as the criterion variable and is most often denoted by the letter "Z". In private enterprise, profit, sales, return on investment, market share, customer satisfaction, and employee turnover are examples of criterion variables commonly of interest to managers. In the public sector, cost, productivity, and quality of service are frequently used criteria. Within the Air Force, costs, operational readiness rates, fill rates, turnaround times, system reliability, and employee satisfaction are familiar criterion variables.

The objective in any decision situation is to adopt a policy or implement a management action that leads to a satisfactory level of (system) performance as measured by the criterion variable. In general, objectives are of two basic types: (1) satisfaction objectives and (2) optimization objectives. The satisfaction objective is one for which the decision taker specifies a value of the criterion variable that is "good enough." If a system manager specifies that the system of interest should have a reliability of at least 0.90, the manager has determined a level or value of the criterion variable (reliability) that is satisfactory or good enough for all conditions in which that system will exist and operate. In other situations, the decision taker's objective is to optimize, i.e., maximize or minimize, the value of the criterion variable subject to a set of conditions, limitations, or constraints that must be satisfied. Maximizing system reliability, subject to constraints of cost, time, and available technology, is a familiar optimization objective. The objective of any LP model is to optimize the value of the criterion variable, subject to a set of constraints or limiting conditions. Using standard LP notation, the objective can be expressed as:

maximize Z, or
minimize Z

Having determined the decision taker's criterion and associated (optimization) objective, the next step in building the LP model is to complete the set of system components. For each candidate variable considered for inclusion in the system, the analyst must pose the question, "Does it significantly affect or influence the criterion variable?" If the answer to this question is "yes" the variable is included. If the answer is "no" the variable is excluded. Obviously, the process of both nominating

and discriminating between relevant and irrelevant variables is very subjective, being greatly affected by the analyst's education, experience, values, insight, creativity, and judgment.

For each component, variable, or constant included in the system model set, the analyst must next ask, "Can the manager influence or change the value of this component, i.e., can the manager do anything about it?" If the answer is "yes," the variable is a resource. If the answer is "no," it is an environmental factor. Resource variables are often referred to as decision variables or management action variables, reflecting the capability of the manager to manipulate or influence these factors. These variables are denoted by the symbol "X_j" where the subscript "j" is an index used to label the variables ($j = 1,2,\ldots,n$). Environmental components in a system generally take one of two basic forms. First, they may be variables which operate more or less independently of the manager's control. For example, the demand for a particular item stocked by Base Supply is a variable over which Base Supply management has no control and is, therefore, part of their environment. The second form in which environmental components may be incorporated into a decision system model is as the domain boundary for a decision on resource variable. Most managers, for example, must conduct their operations and activities within budget or financial constraints imposed by other individuals or agencies. Consequently, while the amount of money to be allocated by the manager between competing alternatives is a decision variable within the manager's control, the total amount to be allocated is not within the manager's control, and is, therefore, an environmental factor. In general, when any variable is bounded in its domain, those limiting variable values define the boundary of the system and its environment.

System Relationships

A system is more than a set or collection of two or more elements. A system is a set of two or more components that are, in some fashion, linked by one or more relationships. The relationships in a system can be described in a number of ways. First, relationships tend to be "directional" in nature, i.e., changes in one variable produce or induce changes in others:

Figure 16-1

In the simplest two-component system illustrated in Figure 16-1, a change in variable X_1, induces some change in variable X_2. In this example, X_2 is said to be dependent (on changes in X_1), while X_1 is independent (of changes in X_2). In functional notation, this expression can be denoted as:

$$X_2 = f(X_1)$$

System relationships can also be characterized as either direct or inverse. In a direct relationship, a variation in the independent variable (i.e., either increase or decrease) results in a change in the dependent variable in the same direction. In an inverse relationship, a variation in a particular direction by the independent variable leads to a change in the opposite direction by the dependent variable. Symbolically, a direct relationship is conventionally denoted by (+) while an inverse relationship is indicated by (-).

Figure 16-2
Direct Relationship

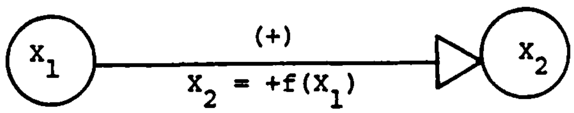
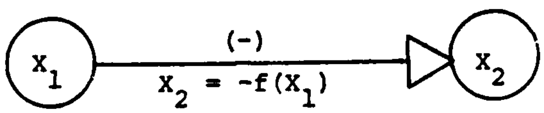


Figure 16-3
Inverse Relationship



To complete the description of a relationship, it is necessary to specify or describe the exact character of the link, ideally, through some type of mathematical expression (equality or inequality) that describes best the behavior of the dependent variable in response to changes of a specified magnitude (and/or rate) in the independent variable. These relationships may be continuous and linear or curvilinear (e.g., exponential, logarithmic, etc.) or they may be less "well-behaved," having skips, spikes, or other irregularities. In LP, as the name implies, all relationships are assumed to be linear. While the assumption of linear relationships is often a gross oversimplification of the real system of interest, it has the advantage of facilitating the development of an analytical solution or policy.

In an LP model, there are two basic types of relationships: (1) the objective or criterion function and (2) system (boundary) constraints. The objective (criterion) function expresses a linear relationship between the dependent criterion variable (Z) and the various independent resource (decision) variables (X_j):

$$\text{Optimize:} \quad Z = C_1 \cdot X_1 + C_2 \cdot X_2 + \ldots + C_n \cdot X_n$$

where C_j is a constant (model parameter) representing the marginal contribution to Z resulting from a unit change in X_j. Expressed more efficiently in summation notation, the objective function can be written as:

$$\text{Optimize:} \quad Z = \sum_{j=1}^{n} \left(C_j \cdot X_j \right)$$

All of the remaining relationships in an LP decision system model are considered as constraints to be satisfied in optimizing the objective function. System constraints typically express the relationship between decision variables in the system and various environmental factors (parameters) representing resource limitations and/or system output requirements. The constraints in an LP model can be either equalities or inequalities. The set of constraints can be expressed mathematically as:

$$\begin{array}{l} A_{11} X_1 + A_{12} X_2 + \ldots + A_{1n} X_n \ (\leq, =, \geq) \ b_1 \\ A_{21} X_1 + A_{22} X_2 + \ldots + A_{2n} X_n \ (\leq, =, \geq) \ b_2 \\ \qquad \vdots \\ A_{m1} X_1 + A_{m2} X_2 + \ldots + A_{mn} X_n \ (\leq, =, \geq) \ b_m \end{array}$$

or, alternatively:

$$\sum_{j=1}^{n} a_{ij} X_j \ (\leq, =, \geq) \ b_i \quad (i = 1,2,\ldots,m)$$

where a_{ij} is a constant (model parameter) representing the marginal consumption of resource b_i or, alternatively, the marginal satisfaction of requirement b_i resulting from a unit increase in decision variable X_j.

While system constraints generally relate decision variables and environmental constraints, they may also specify boundary values on the domains of the respective decision variables. In all LP models, for example, decision variables cannot be negative (you can't allocate negative quantities of a resource). Therefore, LP models must include a set of "nonnegativity" constraints of the form:

$$X_j \geq 0 \ (j = 1,2,\ldots,n)$$

The General LP Model

To summarize, every LP decision system model is a variation of the following general formulation:

Optimize: $Z = C_1 X_1 + C_2 X_2 + \ldots + C_n X_n$

subject to (s.t.):

$$A_{11} X_1 + A_{12} X_2 + \ldots + A_{1n} X_n \ (\leq, =, \geq) \ b_1$$

$$A_{21} X_1 + A_{22} X_2 + \ldots + A_{2n} X_n \ (\leq, =, \geq) \ b_2$$

$$\vdots$$

$$A_{m1} X_1 + A_{m2} X_2 + \ldots + A_{mn} X_n \ (\leq, =, \geq) \ b_m$$

$$X_j \geq 0 \ (j = 1,2,\ldots,n)$$

or, alternatively, in summation notation:

Optimize: $Z = \sum_{j=1}^{n} C_j X_j$

s.t. $\sum_{j=1}^{n} a_{ij} X_j \ (\leq, =, \geq) \ b \quad (i=1,2,\ldots,m)$

$$X_j \geq 0 \ (j=1,2,\ldots,n)$$

Formulation Examples

EXAMPLE 16-1

Problem Statement

You are responsible for inspecting the work of two contractors currently working on the base. From past experience and available historical data, you estimate that each hour spent inspecting Contractor No. 1 will result in a savings to the government of \$75. Similarly, each hour spent inspecting Contractor No. 2 results in an estimated savings of \$50. You estimate that a total of 120 manhours of inspection time are available to allocate between the two contracts each week. Existing policy requires that you devote at least 30% of the available inspection time to each contract. In addition, your transportation budget permits your inspectors to log a total of approximately 600 miles/week in support of their work on these two contracts. You estimate that, on the average, about 4 miles are driven for each hour spent inspecting Contractor No. 1 and 6 miles for each hour spent inspecting Contractor No. 2. How should you allocate the available inspection manhours between the two contracts?

System Analysis

System Components:

1. Criterion Variable (Z): savings in dollars, the objective is to maximize ($Z \geq 0$).

2. <u>Decision Variables</u> (X_j):

X_1 = manhours allocated to inspection of Contractor No. 1 ($X_1 \geq 36$)

X_2 = manhours allocated to inspection of Contractor No. 2 ($X_2 \geq 36$)

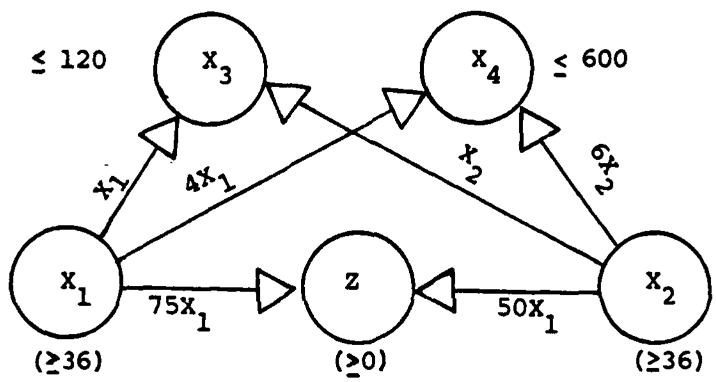
3. <u>Environmental Variables</u> (X_j)

X_3 = total inspection manhours allocated ($X_3 \leq 120$)

X_4 = total miles driven inspecting Contractors 1 and 2 ($X_4 \leq 400$)

<u>System Relationships</u>

<u>Figure 16-4</u>



<u>LP Model</u>

maximize: $Z = 75X_1 + 50X_2$

$$
\begin{array}{llcrl}
\text{s.t.} & X_1 + X_2 & \leq & 120 & (X_3) \\
 & X_1 & \geq & 36 & \\
 & X_2 & \geq & 36 & \\
 & 5X_1 + 8X_2 & \leq & 600 & (X_4)
\end{array}
$$

(Note: the nonnegativity constraints, $X_j \geq 0$, required in all LP models are insured by the policy conditions that require at least 30% of the available inspection time to be allocated to each contract, i.e., at least 120 x 0.3 = 36 manhours).

EXAMPLE 16-2

Problem Statement

You are estimating the cost of 1,000 tons of concrete to be used in a project to repair the airfield pavement at your base. Sand for the concrete is available from two sources. Washed beach sand can be obtained at a delivered cost of $60/ton and contains 4 parts fine sand, 3 parts coarse sand, and 5 parts gravel. Clean river sand can be obtained at a delivered cost of $100/ton and contains 3 parts fine sand, 6 parts coarse sand, and 9 parts gravel. Each batch of concrete must contain 4 parts gravel, 3 parts fine sand, and 3 parts coarse sand. How much sand should be purchased from each source?

System Analysis

System Components:

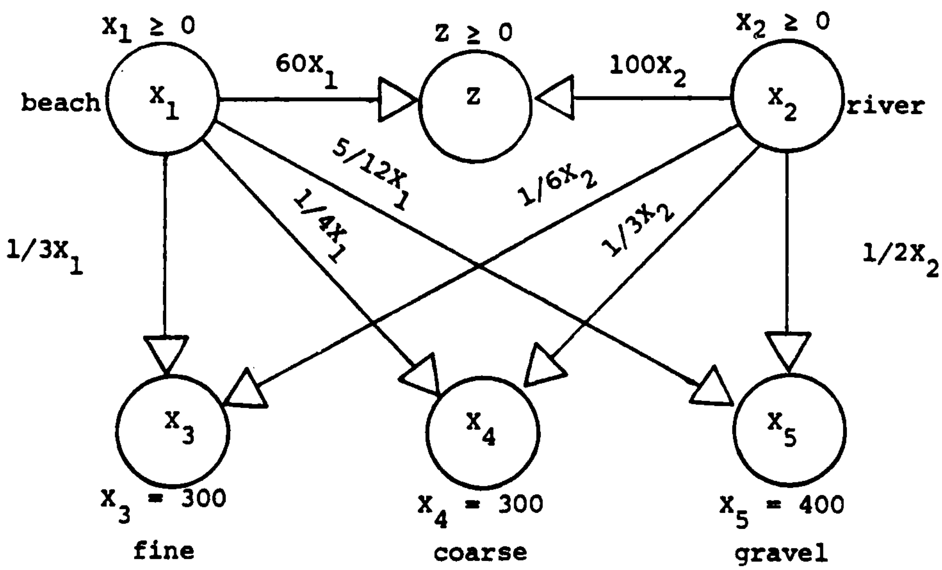
1. Criterion Variable (Z): Cost in dollars; the objective is to minimize Z. ($Z \geq 0$).

2. Decision Variables (X_j):
X_1 = tons of beach sand to purchase ($X_1 \geq 0$)
X_2 = tons of river sand to purchase ($X_2 \geq 0$)

3. Environmental Variables:
X_3 = tons of fine sand used ($X_3 = 3/10 \times 1000 = 300$ tons)
X_4 = tons of coarse sand used ($X_4 = 3/10 \times 1000 = 300$ tons)
X_5 = tons of gravel used ($X_5 = 4/10 \times 1000 = 400$ tons)

System Relationships:

Figure 16-5



LP Model

$$\text{minimize:} \quad Z = 60X_1 + 100X_2$$

$$\begin{aligned} \text{s.t.} \quad 1/3X_1 + 1/6X_2 &= 300 \\ 1/4X_1 + 1/3X_2 &= 300 \\ 5/12X_1 + 1/2X_2 &= 400 \\ X_1 &\geq 0 \\ X_2 &\geq 0 \end{aligned}$$

EXAMPLE 16-3

Problem Statement

A new Federal Environmental Protection Organization is being formed, and there are 2,000 professional and 1,000 nonprofessional positions which need to be filled. Recruiting costs average \$1,000 for each professional position and \$400 for each nonprofessional position. Typically, these costs are 20% higher than average for recruiting women and 30% higher than average for recruiting minorities (men and women). HEW has examined state employment records and has mandated that women should constitute at least 40% of new hirings within the agency and minorities should constitute at least 50% of new hirings. How should the required positions be filled?

System Analysis

System Components:

1. Criterion Variable (Z): Cost in dollars; the objective is to minimize Z ($Z \geq 0$).

2. Decision Variables (X_{ijk}): Let X_{ijk} = the number of individuals of type (i, j, k) hired where:

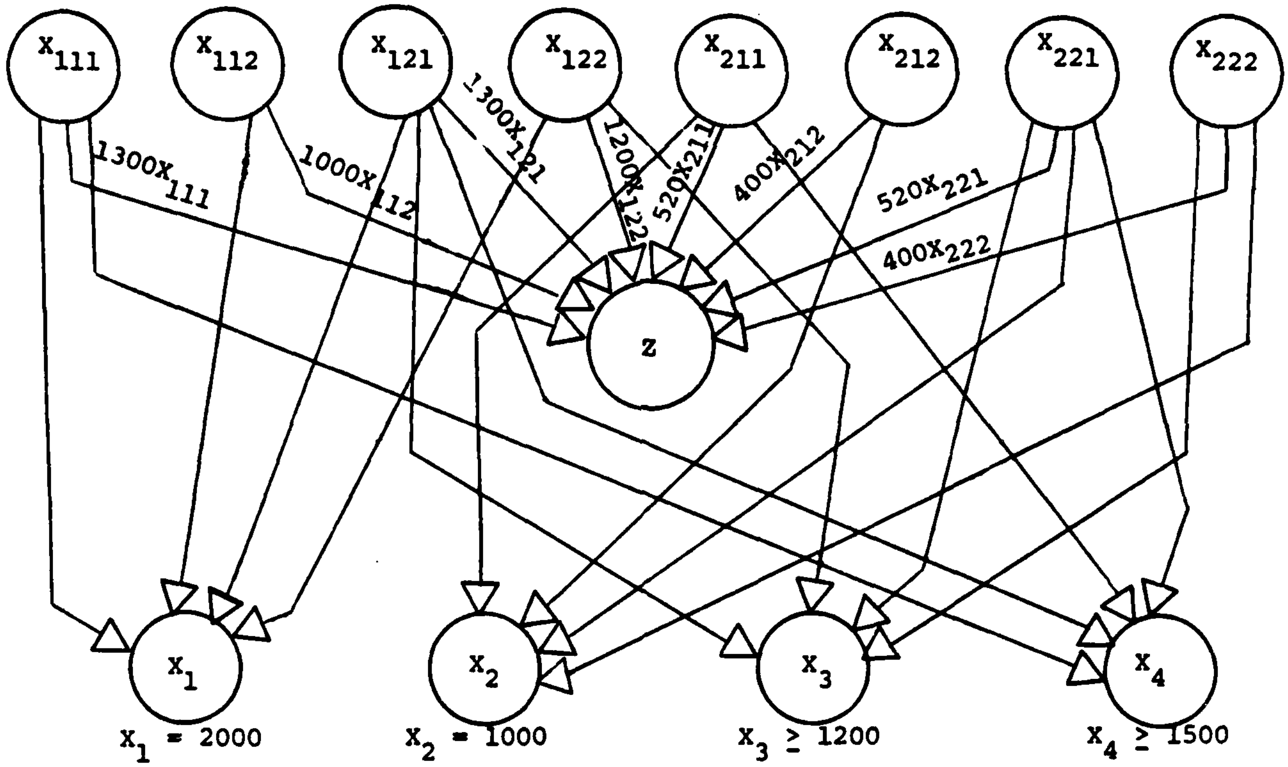
   i = 1: professional
   = 2: nonprofessional
   j = 1: men
   = 2: women
   k = 1: minority
   = 2: nonminority

3. Environmental Variables (X_j):

   X_1 = total no. of professionals hired ($X_1 = 2000$)
   X_2 = total no. of nonprofessionals hired ($X_2 = 1000$)
   X_3 = total no. of women hired ($X_3 \geq 1200$)
   X_4 = total no. of minorities hired ($X_4 \geq 1500$)

System Relationships

Figure 16-6



LP Model

$$
\begin{aligned}
\text{minimize:}\quad Z = {} & 1300X_{111} + 1000X_{112} + 1300X_{121} \\
& + 1200X_{122} + 520X_{211} + 400X_{212} + 520X_{221} + 480X_{222} \\
\text{s.t.}\quad & X_{111} + X_{112} + X_{121} + X_{122} = 2000 \\
& X_{211} + X_{212} + X_{221} + X_{222} = 1000 \\
& X_{121} + X_{122} + X_{221} + X_{222} \geq 1200 \\
& X_{111} + X_{121} + X_{211} + X_{221} \geq 1500 \\
& X_{ijk} \geq 0 \quad (i,\ j,\ k = 1,2)
\end{aligned}
$$

Note: Solution values for the decision variables in this problem would be large enough to mitigate the effects of rounding off a fractional answer, i.e., the divisibility assumption of the LP model should pose no practical problem or limitation.

EXAMPLE 16-4

Problem Statement

You are responsible for contracting for the delivery of a particular type of aviation fuel for each of four installations. The fuel can be purchased from three sources. The monthly fuel requirements (in railroad tank car lots) for each base, monthly delivery capacity for each vendor, and the associated costs are summarized in the following table:

	INSTALLATION				
VENDOR	1	2	3	4	CAPACITY
1 2 3	\$5500 \$8000 \$6000	\$6500 \$6500 \$8000	\$7000 \$6000 \$9000	\$6750 \$7000 \$6000	8 10 12
DEMAND	10	5	8	7	30

How much fuel should be purchased from each vendor for delivery to the respective bases?

System Analysis

System Components:

1. Criterion Variable (Z): Cost in dollars; the objective is to minimize Z ($Z \geq 0$).

2. Decision Variables: Let X_{ij} = the number of tank cars of fuel shipped from vendor i to installation j, where i = 1,2,3 and j = 1,2,3,4.

3. Environmental Variables:

X_1 = demand at installation 1 ($X_1 = 10$)

X_2 = demand at installation 2 ($X_2 = 5$)

X_3 = demand at installation 3 ($X_3 = 8$)

X_4 = demand at installation 4 ($X_4 = 7$)

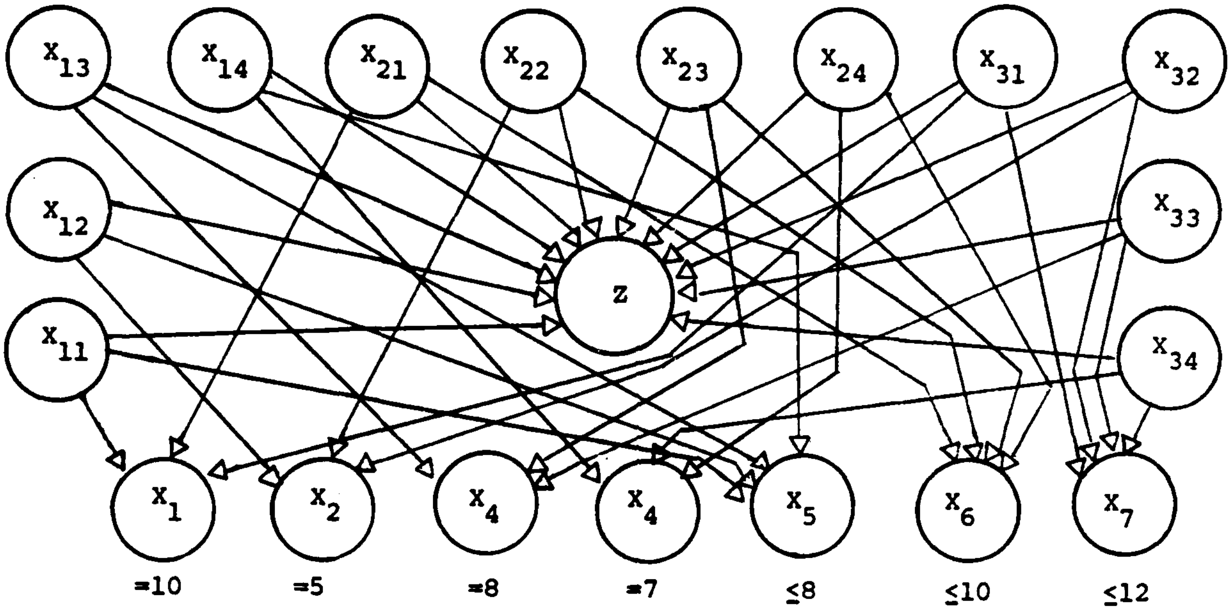
X_5 = supply capacity at vendor 1 ($X_5 \leq 8$)

X_6 = supply capacity at vendor 2 ($X_6 \leq 10$)

X_7 = supply capacity at vendor 3 ($X_7 \leq 12$)

System Relationships

Figure 16-7



LP Model

minimize: $Z = 5500X_{11} + 6500X_{12} + 7000X_{13} + 6750X_{14}$
$+ 8000X_{21} + 6500X_{22} + 6000X_{23} + 7000X_{24}$
$+ 6000X_{31} + 8000X_{32} + 9000X_{33} + 6000X_{34}$

s.t.

$$
\begin{aligned}
X_{11} + X_{21} + X_{31} &= 10 \\
X_{12} + X_{22} + X_{32} &= 5 \\
X_{13} + X_{23} + X_{33} &= 8 \\
X_{14} + X_{24} + X_{34} &= 7 \\
X_{11} + X_{12} + X_{13} + X_{14} &\leq 8 \\
X_{21} + X_{22} + X_{23} + X_{24} &\leq 10 \\
X_{31} + X_{32} + X_{33} + X_{34} &\leq 12 \\
X_{ijk} \geq 0 \quad (i = 1,2,3;\ j &= 1,2,3,4)
\end{aligned}
$$

SOLUTION TECHNIQUES

One of the primary factors leading to the popularity of LP as an aid to decision taking is that solution algorithms are readily available. For very simple systems involving only two or three decision variables, the optimal policy or solution can be determined graphically. For more complex problems, a solution procedure-the simplex method-offers a relatively efficient technique for determining the optimal policy. While the graphical solution is not often used in practice because of its restricted applicability, it provides a good basis for understanding the manner in which more complex problems can be solved.

The Graphical Method

Consider again the LP model developed in Example 16-1:

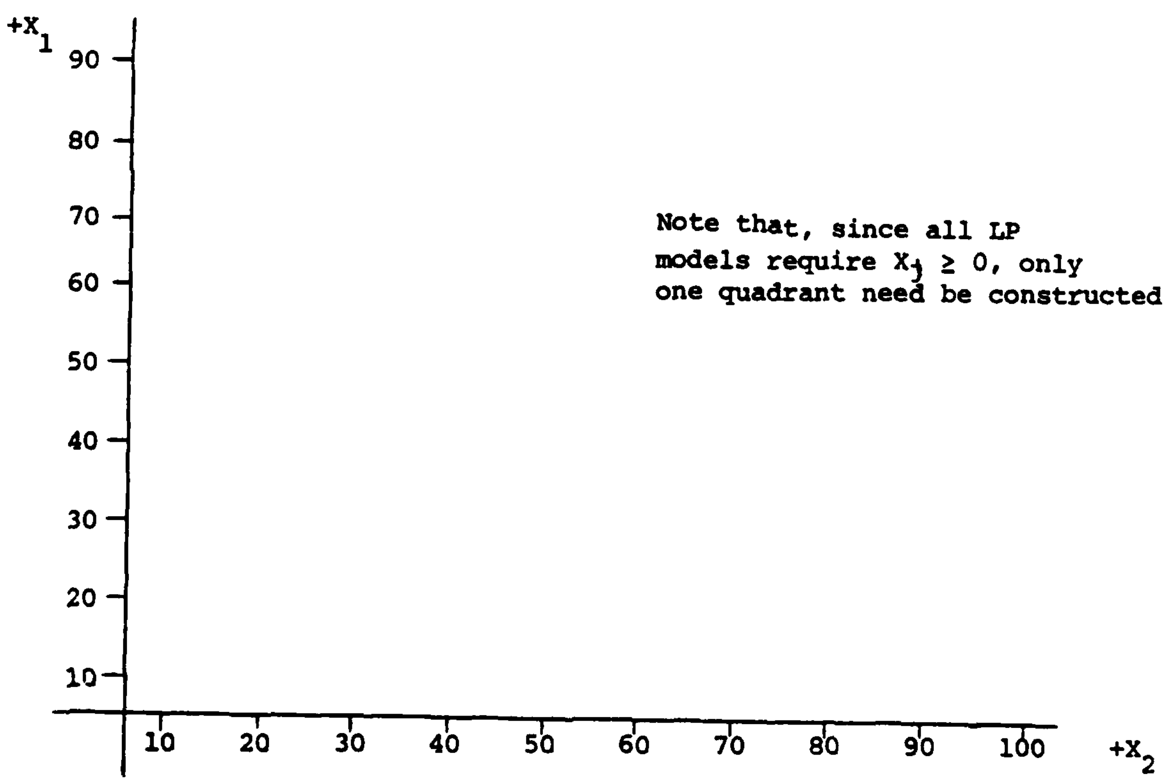
$$\text{maximize } Z = 75X_1 + 50X_2$$

$$\begin{array}{llcl} \text{s.t.} & X_1 + X_2 & \leq & 120 \\ & X_1 & \geq & 36 \\ & X_2 & \geq & 36 \\ & 5X_1 + 8X_2 & \leq & 400 \end{array}$$

The first step in the graphical solution procedure is to construct an orthogonal coordinate system with each of the axes representing one of the decision (resource) variables:

System Relationships

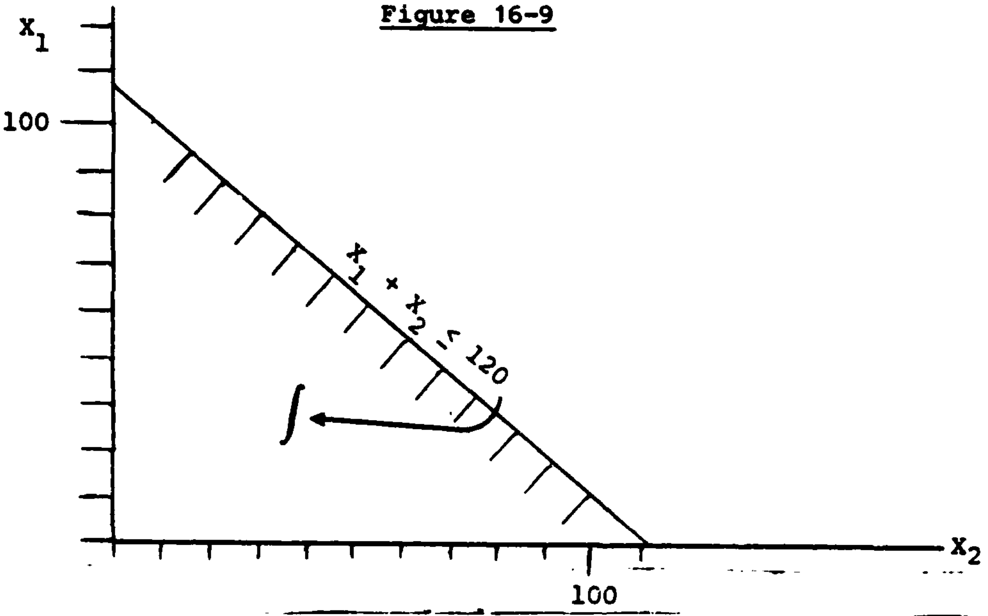
Figure 16-8



Note that, since all LP models require $X_j \geq 0$, only one quadrant need be constructed.

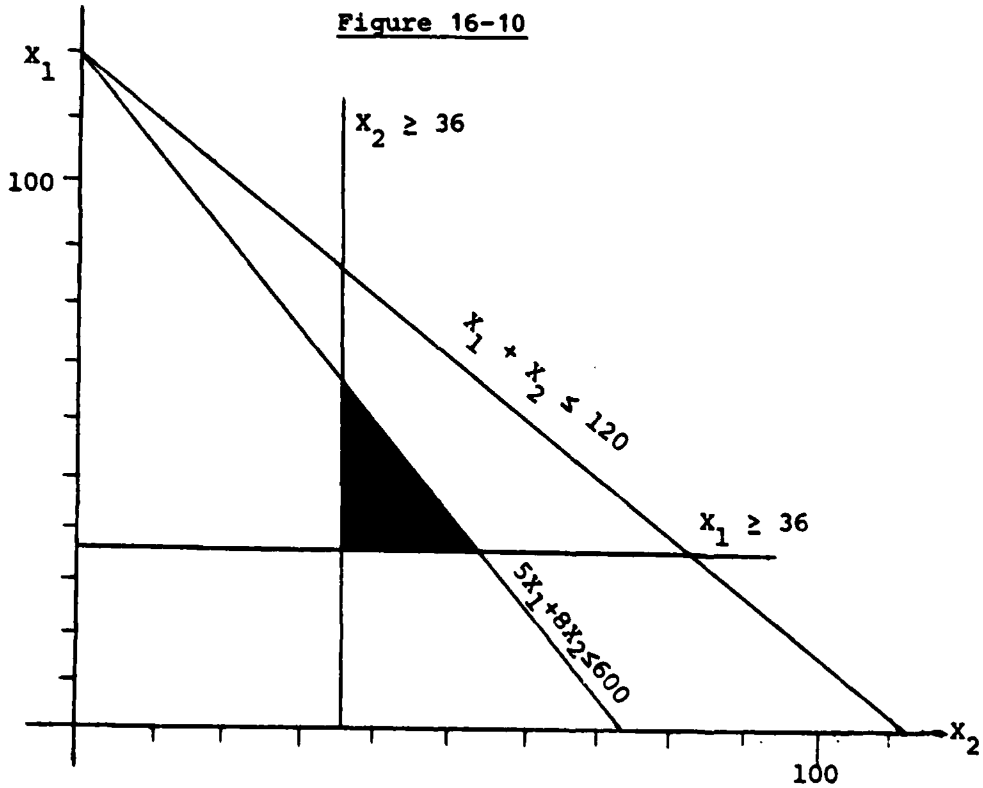
Next, each of the constraints is constructed in this coordinate system:

Figure 16-9



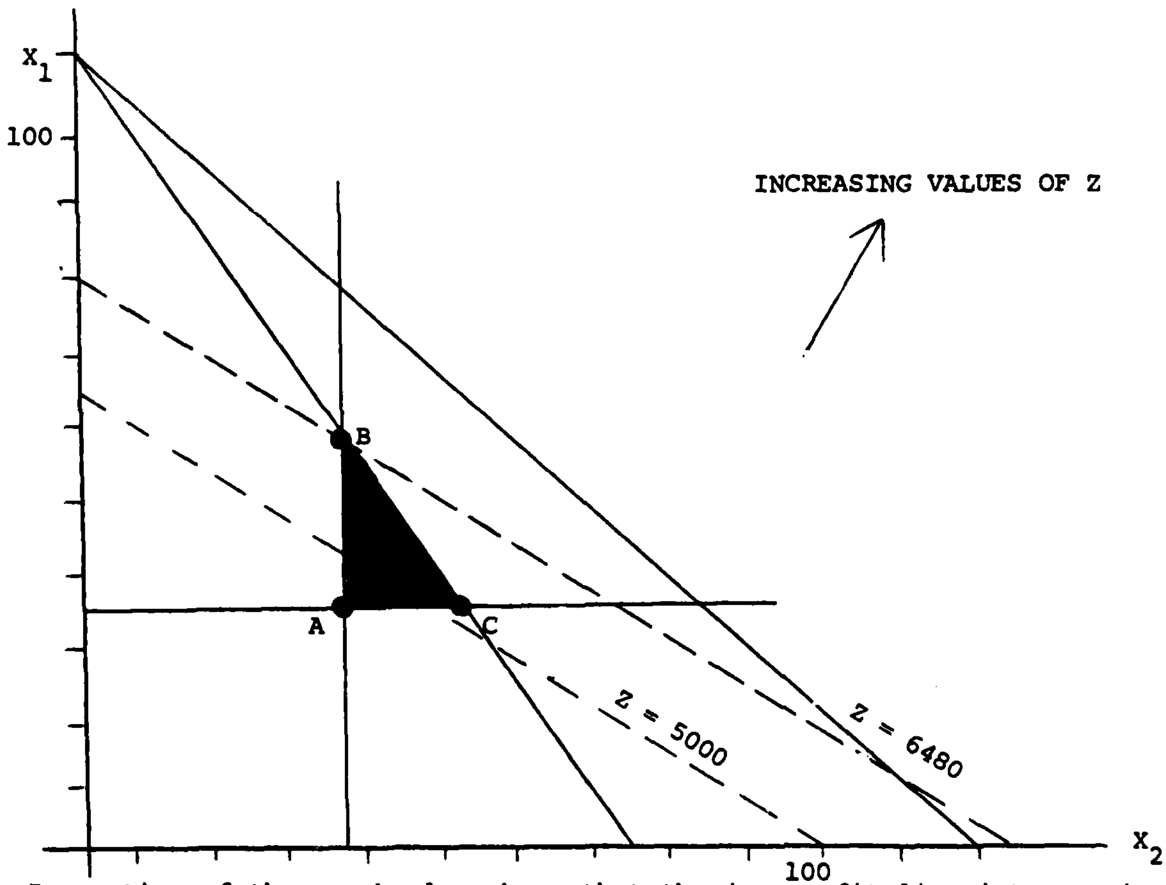
Constructing the first constraint, $X_1 + X_2 \leq 120$, defines a triangular region bounded by that constraint and each of the axes. In a similar fashion, the other constraints can be constructed in the coordinate system:

Figure 16-10



The shaded area represents that region of points or coordinates which simultaneously satisfy all of the constraints. Because only the points in the shaded area represent feasible solutions to the problem, this area is referred to as the feasible region or (feasible) solution space. Note, too, that for this particular problem, the first constraint constructed, i.e., $X_1 + X_2 \leq 120$, does not bound the feasible region at any point - it is "redundant" and is not, in this case, really a limitation. But which of the infinite number of points in the feasible solution space (X_1, X_2) yields the greatest value of Z (remember, the objective is to maximize Z)? The answer is provided by constructing a line on the graph representing the objective function. This can be done by arbitrarily selecting a value for Z, e.g., $Z = 5000 = 75X_1 + 50X_2$, and drawing that line on the graph as shown in Figure 16-11. Any value of Z selected will yield an "iso-profit" line parallel to the first. Since the objective here is to maximize Z, the problem is to find the iso-profit line corresponding to the largest value of Z, while still including a point in the feasible region. By inspection, it is apparent that as iso-profit lines are constructed in the direction away from the origin, the value of Z increases, and vice versa.

Figure 16-11



Inspection of the graph also shows that the iso-profit line intersecting point B will result in the greatest value of Z. Point B represents the intersection of the constraint lines:

$$X_2 = 36$$

$$5X_1 + 8X_2 = 600$$

solving: $5X_1 + 8(36) = 600$

$$X_1 = 312/5 = 62.4$$

Therefore, the optimal feasible solution to this problem is $X_1 = 62.4$, $X_2 = 36$, and $Z = 75(62.4) + 50(36) = 4680 + 1800 = \$6480$. The fact that the optimal solution fell at a corner point of the feasible region is, of course, not merely a coincidence. It is characteristic of all LP models. A set of constraints which are expressed as linear inequalities must necessarily form a convex set. Further, given a linear objective function, regardless of its slope of the direction of optimization, the optimal solution will always include a corner (extreme) point of a linear convex set. This characteristic holds whether we are dealing with two, three, or n decision variables/dimensions. This characteristic is the key to the systematic solution of all LP models because it means one need investigate only a finite number of corner points to be assured of finding an optimal feasible solution if one exists. Since corner points are, in effect, defined by the intersection or two or more constraints, the systematic determination of an optimal solution is mathematically equivalent to the iterative solution of sets of simultaneous linear equations defining the respective corner points of the feasible region.

Aberrations in LP Models

In building and solving LP models of various decision systems, it is possible to encounter a number of aberrations to the general LP model and its solution. Three such contingencies are commonly encountered: (1) alternative optimal solutions, (2) unbounded solutions, and (3) nonexisting feasible solutions.

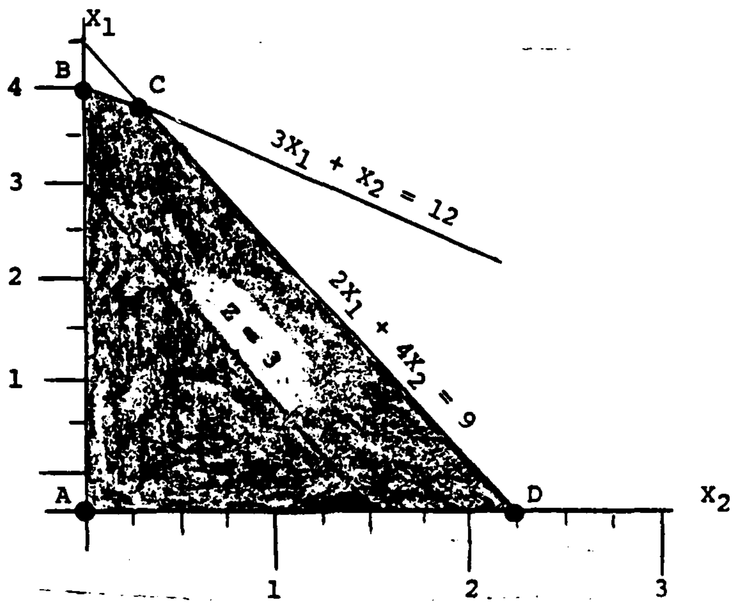
Alternate Optimal Solutions

Consider the following LP decision system model:

$$\begin{aligned} \text{maximize:} \quad & Z = X_1 + 2X_2 \\ \text{s.t.:} \quad & 2X_1 + 4X_2 \leq 9 \\ & 3X_1 + X_2 \leq 12 \\ & X_1, \; X_2 \geq 0 \end{aligned}$$

A graph of this model is illustrated in Figure 16-12:

Figure 16-12



Note that the objective function for Z = 3 is parallel to the constraint $2X_1 + 4X_2 \leq 9$. It is apparent that when the objective function line is (graphically) moved away from the origin in an attempt to maximize Z, points C and D will be encountered simultaneously, as will the infinite number of points defining the line connecting these points. In fact, the optimal solution occurs when the objective function is coincident with line $2X_1 + 4X_2 = 9$. Any point on the line segment connecting points C and D will yield the same optimal value of Z (in this case, Z = 4.5). Whenever the objective function is parallel to a binding constraint, an infinite number of alternate optima exist.

Unbounded Solutions

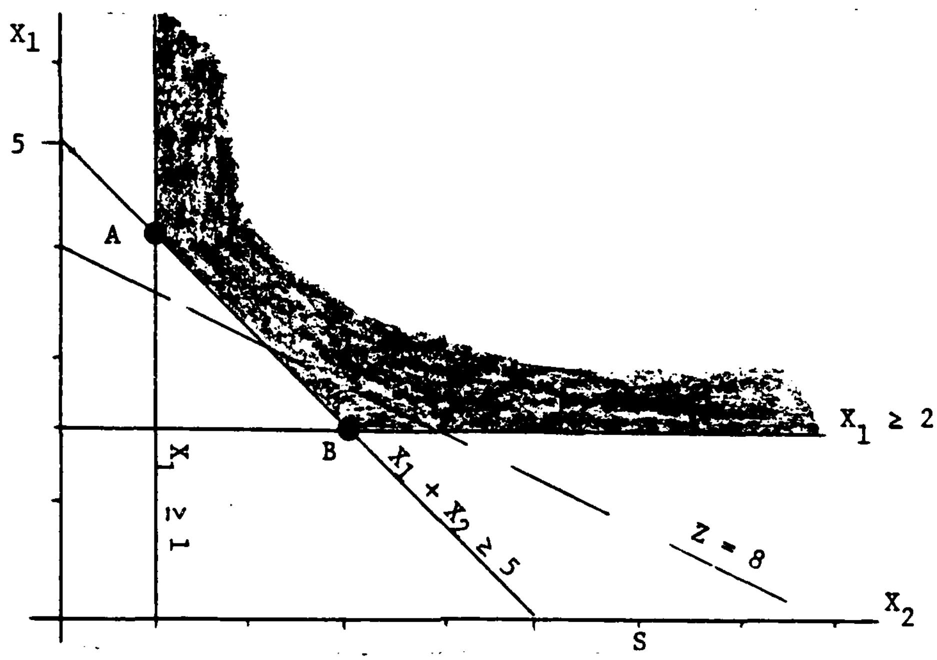
Consider the following LP decision system model:

$$\text{optimize:} \quad Z = 2X_1 + X_2$$

$$\text{s.t.} \quad \begin{aligned} X_1 + X_2 &\geq 5 \\ X_1 \quad\quad &\geq 2 \\ X_2 &\geq 1 \end{aligned}$$

Figure 16-13 graphically illustrates this model:

Figure 16-13



Note that the objective function line with Z=8 has been included as a reference. If the optimization objective for this problem was to minimize Z, it is apparent that the optimal feasible solution will be at point B, the extreme (corner) point of the solution space that is closest to the origin (X_1 = 2, X_2 = 3, Z = 7). Suppose, however, that the stated objective was to maximize Z. Since the feasible region is unbounded (in this case) in the direction away from the origin, there is no limit to the distance away from the origin at which objective function lines can be constructed. Mathematically, Z can be increased without bound. When this situation is encountered, it indicates that the decision system model has been formulated incorrectly. Go back and check your analysis and logic.

No Feasible Solution

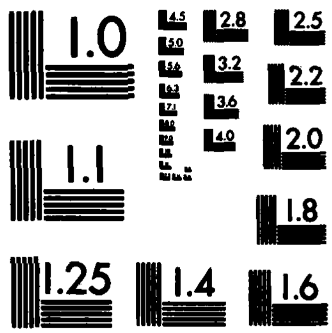
In many instances, a decision situation may be described for which no feasible solution exists, i.e., it is not possible to simultaneously satisfy all of the conditions or constraints on the problem. This is particularly true for systems having large numbers of constraints - decision systems that are "tightly bound". Consider the following example:

$$\text{optimize:} \quad Z = 2X_1 + X_2$$

$$\begin{aligned} \text{s.t.} \quad X_1 + X_2 &\leq 4 \\ X_1 \quad\quad &\geq 5 \\ X_2 &\geq 6 \end{aligned}$$

AD-A121 990 QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS J E ENGEL APR 80 AFIT-LS-32 5/7

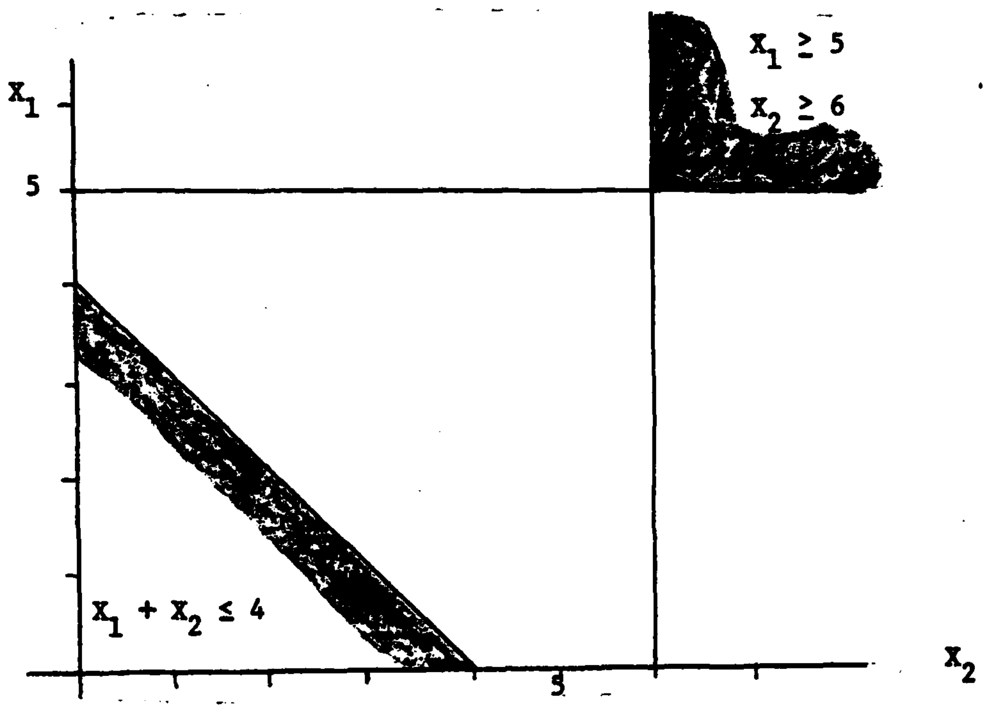
UNCLASSIFIED F/G 15/5 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Figure 16-14 illustrates the solution space for this model:

Figure 16-14



It is apparent that there is no point that can, at the same time, meet all three conditions. In this case, there is no solution to the decision situation as it has been described in the model.

THE SIMPLEX METHOD

While the graphical solution procedure works well for very simple problems having only two decision variables, there are few realistic problems that can be adequately modeled in two dimensions. For problems involving hundreds of variables and constraints, a more powerful and efficient technique is required. This technique is the simplex method. The simplex method is a straightforward procedure that begins with an initial feasible solution (i.e., a corner point of the feasible region) and sequentially proceeds toward the optimal solution by examining adjacent points in the solution space. The procedure iterates or moves toward the optimal solution in such a way that each new solution examined is better than the previous one. When a solution is examined and found to be less optimal than the previously investigated corner point, the procedure terminates. The simplex algorithm procedes through the following basic steps:

(1) Determine an initial basic feasible solution;
(2) Check the solution for optimality:
a. if optimal, stop;
b. if not optimal, continue;
(3) Locate the adjacent corner point of the solution space that makes the greatest contribution to optimizing Z;
(4) Go to step (2).

Mechanically or, more appropriately, mathematically, the simplex proceeds by solving the subset of simultaneous linear equations which define a particular corner point or solution. If the solution obtained is not optimal, a new subset of constraint equations, corresponding to one of the adjacent extreme points, is systematically selected and solved. The iterative process continues until the optimal solution is determined.

Standard (Canonical) Form of the LP Model

The first step in the simplex solution procedure is to convert the LP model resulting from your analysis of the decision system of interest to a standard or canonical form having the following characteristics: (1) all structural (nonnegativity) constraints are expressed as equalities; (2) the right-hand side constants in the constraint set are all nonnegative; (3) all variables are nonnegative; (4) the objective is maximization. Converting the LP model to a standard form greatly facilitates the solution procedure, particularly when being accomplished manually, i.e., without one of the many available LP computer codes. In summation notation, the standard form of the LP model is:

$$\text{maximize:} \quad Z = \sum_{j=1}^{n} c_j x_j$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij} x_j = b_i \ (i = 1,2,\ldots,m)$$

$$X_j \geq 0 \ (j = 1,2,\ldots,n)$$

Conversion of an LP Model to Standard Form

In general, the LP model resulting from analysis of the decision system will have mixed constraints, i.e., equalities and inequalities ($\leq$, $\geq$, or both).

Conversion of $\leq$ Constraints.

For each $\leq$ constraint i, introduce a slack variable (S_i) to the left side of the expression and convert the inequality to an equality. For example, given:

$$\text{maximize:} \quad Z = 5X_1 + 8X_2$$

$$\begin{aligned} \text{s.t.} \quad 3X_1 + 4X_2 &\leq 12 \\ 2X_1 + 3X_2 &\leq 6 \\ X_1, X_2 &\geq 0 \end{aligned}$$

converting to standard form, this becomes:

$$\begin{array}{lll} \text{maximize:} & Z = 5X_1 + 8X_2 & \\ \text{s.t.} & 3X_1 + 4X_2 + S_1 & = 12 \\ & 2X_1 + 3X_2 + S_2 & = 6 \\ & X_1, X_2, S_1, S_2 & \geq 0 \end{array}$$

Since the left side of the constraint must be ≤ the right side constant (b_i), the variable S_i represents the "slack" between the left and right sides of the equality. In physical terms, if b_i represents the total amount available of a particular resource, S_i represents the unused portion of that total. Of course, when all of the available resource i (b_i) is used and the constraint is "binding", S_i=0. Note that slack variables do not appear in the objective function or, alternatively, they appear in the objective function with a coefficient (c_j) equal to zero.

Conversion of ≥ Constraints

For each ≥ inequality constraint, the left side must be at least equal to the right side value b_i. It can be greater, but it cannot be less. The ≥ inequality is converted to an equality by subtracting a surplus variable (E_i) from the left side. In addition, it is also necessary to introduce another variable, called an artificial variable (A_i). The artificial variable is necessary to provide a starting point in the solution process, i.e., an initial basic feasible solution. The rationale for introducing the artificial variable will be further discussed shortly. As an example:

$$\begin{array}{lll} \text{maximize:} & Z = 5X_1 + 8X_2 & \\ \text{s.t.} & 3X_1 + 4X_2 & \geq 12 \\ & 2X_1 + 4X_2 & \geq 6 \\ & X_1 + X_2 & \geq 0 \end{array}$$

converting to standard form:

$$\begin{array}{lll} \text{maximize:} & Z = 5X_1 + 8X_2 & \\ \text{s.t.} & 3X_1 + 4X_2 - E_1 + A_1 & = 12 \\ & 2X_1 + 3X_2 - E_2 + A_2 & = 6 \\ & X_1, X_2, E_1, E_2, A_1, A_2 & \geq 0 \end{array}$$

Unfortunately, there is no guarantee that one or more of the artificial variables won't show up in the final solution (at a nonzero level). A number of "tricks" have been devised to preclude artificial variables in the final solution. The most popular of these is the method of penalties - more commonly referred to as the "Big M" method. Since the objective is to maximize, introducing A_i in the objective function with a very large $-C_j$ coefficiant (represented by "M") insures that if any A_i shows up in solution, a very large penalty in terms of the objective is incurred. Therefore, the standard form of the example is:

$$\begin{array}{lll} \text{maximize:} & Z = 5X_1 + 8X_2 - MA_1 - MA_2 & \\ \text{s.t.} & 3X_1 + 4X_2 - E_1 + A_1 & = 12 \\ & 2X_1 + 3X_2 - E_2 + A_2 & = 6 \\ & X_1, X_2, E_1, E_2, A_1, E_2 & \geq 0 \end{array}$$

Equality Constraints

Many LP models have equality constraints as well as those inequalities just discussed. On the surface, it would seem that no action is required here to convert it to standard form. Unfortunately, such is not the case. For purposes of finding a starting point (as in the case of the $\geq$ inequality), it is necessary to introduce an artificial variable (A_i) in each equality constraint and a $-MA_i$ term in the objective function. To summarize, then, consider the following LP model:

$$\text{maximize:} \quad Z = 2X_1 + 5X_2$$

$$\begin{aligned} \text{s.t.} \qquad X_1 + 3X_2 &\leq 30 \\ 4X_1 + 2X_2 &\geq 15 \\ X_1 + X_2 &= 10 \\ X_1, X_2 &\geq 0 \end{aligned}$$

converting to standard form, this becomes:

$$\text{maximize:} \quad Z = 2X_1 + 5X_2 - MA_2 - MA_3$$

$$\begin{aligned} \text{s.t.} \qquad X_1 + 3X_2 + S_1 &= 30 \\ 4X_1 + 2X_2 - E_2 + A_2 &= 15 \\ X_1 + X_2 + A_3 &= 10 \\ X_1, X_2, E_2, A_2, A_3 &\geq 0 \end{aligned}$$

Conversion of Negative Right Side Values

Recall that in its standard form, negative b_i values are not permitted. Should a $-b_i$ value result from the decision system analysis, simple multiply both sides of the constraint by -1. If the constraint is an inequality, multiplying through by -1 reverses the sense or direction of the inequality. Once the $-b_i$ value has been converted to a $+b_i$, we proceed as before, introducing, as necessary, slack, surplus, and/or artificial variables. Consider, for example, the constraint:

$$X_1 + X_2 \geq -5$$

to convert to standard form, multiply both sides by -1:

$$-X_1 - X_2 \leq 5$$

and add a slack variable to convert the inequality to an equality:

$$-X_1 - X_2 + S_1 = 5$$

Conversion of Minimization Objectives

While it is not absolutely necessary to convert minimization problems to maximize problems, it is convenient to do so because only one set of procedural rules need be mastered. Here, too, the conversion procedure is

straightforward: simply multiply by -1. In other words:

$$\text{minimizing } Z = \sum_{j=1}^{n} C_j X_j \quad \text{is equivalent to} \quad \text{maximizing } (-Z) = \sum_{j=1}^{n} (-C_j) X_j$$

The Solution Procedure

Converting the LP model to standard form effectively results in a system of simultaneous linear equations. In general, in any such system having m equations and n unknowns, there can be a unique solution, an infinite number of solutions, or no solution, depending on the relative size of m and n. When $n > m$, any m variables can be expressed in terms of the remaining (n-m) variables. Consequently, there are, in this case, an infinite number of solutions. This is essentially the situation when an LP model has been converted to standard form. If the original model had n (decision) variables and m constraints, when converted to standard form, at least one slack (S_i), surplus (E_i), and/or artificial variable is introduced in each of the m constraints. Consequently, the standard form will have at least (m+n) variables and m constraint equations. However, a unique solution can be obtained by setting any n unknown variables to zero and solving for the remaining m unknown variables. At the maximum, i.e., through exhaustive explicit enumeration, the total number of solutions to be examined corresponds to the total number of ways of selecting m variables from (m+n) variables. This is given by:

$$C_m^{(n+m)} = \frac{(n+m)!}{(n+m-m)!m!} = \frac{(n+m)!}{n!m!}$$

Consider again Example 16-4 in which the resulting LP model had 12 variables and 7 constraints:

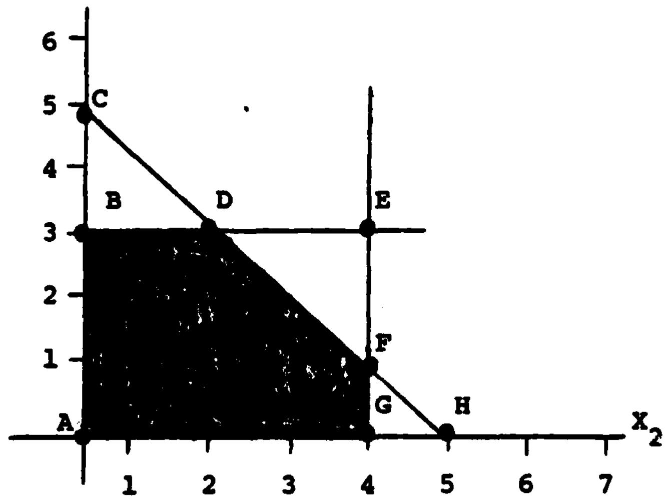
$$C_{12}^{(12+7)} = \frac{19!}{12!7!} = 50,388$$

Fortunately, of this total, we need only look at the feasible subset in which $X_j \geq 0$. Since, as we have seen, the optimal solution will be at one of the corners or extreme points of this feasible region, the problem is actually much more manageable than it might seem. Because the simplex procedure is predicated on these structural characteristics (and because efficient simplex-based computer codes have been developed), it can rather efficiently solve problems having hundreds of variables and constraints. Before briefly describing the simplex algorithm, it will be convenient to introduce a few basic terms:

<u>Basic Solution</u>: Any point defined by the intersection of two (or more) constraint equalities. Basic solutions can be feasible or infeasible. For example, in Figure 16-15, points A, B, D, F, and G are <u>Basic Feasible Solutions</u>. Points C, E, and H are <u>Basic Infeasible Solutions</u>.

<u>Solution Basis</u>: The m variables selected for a solution are said to constitute a (solution) <u>basis</u>.

Figure 16-15



These m variables are referred to as basic variables while the remaining n variables, arbitrarily set equal to zero, are referred to as nonbasic variables.

Degeneracy: In a basic feasible solution, all of the m basic variables are ≥ 0. The n nonbasic variables are, of course, set equal to zero. However, in some cases, one or more of the basic variables may be equal to zero. When this occurs, the basic feasible solution is said to be degenerate.

The next step in the process is to convert the standardized LP model to a table or tabular form in order to facilitate the mechanics of the simplex procedure. To illustrate, consider the following simple example:

minimize: $Z = 3X_1 + X_2$

s.t.

$$X_1 + X_2 = 3 \qquad (1)$$
$$X_1 \leq 5 \qquad (2)$$
$$X_2 \geq 1 \qquad (3)$$
$$X_1, \; X_2 \geq 0$$

converting to standard form:

maximize: $(-Z) = -3X_1 - X_2 - MA_1 + OS_2 + OE_3 - MA_3$

s.t.

$$X_1 + X_2 + A_1 = 3$$
$$X_1 + S_2 = 5$$
$$X_2 - E_3 + A_3 = 1$$
$$X_1, X_2, A_1, S_2, E_3, A_3 \geq 0$$

this model can be expressed as the following linear system:

$$-1Z + 3X_1 + X_2 + MA_1 - 0S_2 - 0E_3 + MA_3 = 0$$

$$0Z + 1X_1 + 1X_2 + 1A_1 + 0S_2 + 0E_3 + 0A_3 = 3$$

$$0Z + 1X_1 + 0X_2 + 0A_1 + 1S_2 + 0E_3 + 0A_3 = 5$$

$$0Z + 0X_1 + 1X_2 + 0A_1 + 0S_2 - 1E_3 + 1A_3 = 1$$

This same system can be somewhat more efficiently described using the following table:

Table 16-1

ROW	BASIS	Z	X_1	X_2	A_1	S_2	E_3	A_3	b_i	r_i
(0)	Z	-1	3	1	M	0	0	M	0	
(1)		0	1	1	1	0	0	0	3	
(2)	S_2	0	1	0	0	1	0	0	5	
(3)		0	0	1	0	0	-1	1	1	

Having constructed this initial table, the procedure continues as follows:

(1) <u>Select a convenient starting basis of m variables which is feasible (find an initial basic feasible solution)</u>.

The basic variables are those which form an identity matrix, i.e., those variables, in addition to Z (which is always in the basis), for which the column vector contains a single "+1" entry and 0 in the remaining positions. At this point, only S_2 satisfies the identity condition. However, through appropriate algebraic manipulation, the "M" entries can be cleared out of the columns corresponding to variables A_1 and A_2, thereby providing an initial basic feasible solution. Specifically, if each entry in row (1) is multiplied by -M and added to the respective values in row (0) (i.e., the elimination-substitution technique), the following matrix results:

Table 16-2

ROW	BASIS	Z	X_1	X_2	A_1	S_2	E_3	A_3	b_i	r_i
(0)	Z	-1	3-M	1-M	0	0	0	M	-3M	
(1)	A_1	0	1	1	1	0	0	0	3	
(2)	S_2	0	1	0	0	1	0	0	5	
(3)		0	0	1	0	0	-1	1	1	

Similarly, to clear the zero from row (0), column A_3, we multiply row (3) by -M and add it to row (0):

Table 16-3

ROW	BASIS	Z	X_1	X_2	A_1	S_2	E_3	A_3	b_i	r_i
(0)	Z	-1	3-M	1-2M	0	0	M	0	-4M	
(1)	A_1	0	1	1	1	0	0	0	3	3/1 = 3
(2)	S_2	0	1	0	0	1	0	0	5	-
(3)	A_3	0	0	1	0	0	-1	1	1	1/1 = 1

This matrix now provides an initial basic feasible solution which can be read directly from the table: $A_1 = 3$, $S_2 = 5$, $A_3 = 1$, and $Z = -4M$. The other variables are, at this point, nonbasic and set equal to zero.

Looking back at Table 16-1, it becomes somewhat easier to see why, in converting the LP model to standard form, it is necessary to include an artificial variable for each equality (=) and (≥) inequality constraint. Had A_1 and A_3 not been introduced here, we would not have been able to readily determine an initial basic feasible solution - we would not have had an identity matrix.

(2) <u>Check the solution for optimality</u>

The next step in the simplex algorithm is to determine if the solution just obtained is optimal. This can be done by inspecting the table. The values in row (0) represent the marginal contribution to the objective when a unit of the respective nonbasic variable is brought into the solution basis. Remember, the basic variables will have a "0" in this position. When all of the values in row (0) are ≥ 0, no further improvement is possible and the optimal solution has been found. Recall that in the process of converting the standard form of the LP model to the tabular form, the signs of the coefficients in the objective function were reversed. Consequently, the test for optimality reflects the condition of all objective function coefficients being ≤ 0 when expressed in standard form. Since the objective is to maximize Z, introducing any units of a variable for which the coefficient is ≤ 0 cannot increase Z and may very likely decrease it. To reiterate, the solution to the problem is optimal <u>when all values in row (0) of the table are ≥ 0</u>. In this example, both X_1 and X_2 have very large negative coefficients. Consequently, this initial basic feasible solution is not optimal.

(3) <u>Select the entering variable</u>

To proceed toward the optimal solution as efficiently as possible, the simplex procedure moves from the current solution point to the adjacent extreme point in the feasible region that results in the greatest contribution to the objective, i.e., to the greatest increase in Z. Procedurally, this is accomplished by bringing one of the nonbasic variables into the solution basis. In doing so, one of the basic variables is driven out. The entering variable is that nonbasic variable having the <u>most negative</u> coefficient in row (0) of the table. Looking at Table 16-3, X_1 has a coefficient of (3-M) while X_2 has a coefficient of (1-2M). Therefore, X_2 becomes the entering variable (because -2M is more negative than -M).

(4) <u>Select the leaving variable</u>

As X_2 enters the solution basis, which variable is driven out (i.e., A_1, S_2, or A_3)? Looking at the X_2 (column) vector in Table 16-3 for each unit X_2 is increased, A_1 is decreased by 1, S_2 is decreased by 0, and A_3 is decreased by 1. The relevant question then becomes how many units the entering variable X_2 can be increased before either A_1, S_2, or A_3 is

exhausted and, therefore, eliminated from the solution basis. The answer, of course, depends on the amount of each basic variable (b_i) and the relative rate at which it is consumed. By calculating the ratio $r_i = b_i/a_{ij}^*$, where a_{ij}^* is the positive (≥ 0) constraint coefficient for the entering variable j and basic variable i, the relative rate at which the various basic variables are driven out of the basis can be determined. These ratios appear in the right end column of the table. The smallest of these r_i values identifies the basic variable driven to zero first as the entering variable is increased. In this example, X_2 can be increased by only 1 before A_3 is driven to zero. Consequently, A_3 is the leaving variable.

(5) Modify the table to reflect the new solution

To establish the new solution basis, we algebraically manipulate the linear system so that the entering variable, in this case X_2, is included in the identity matrix. Specifically, the table is manipulated in such a way as to result in a "+1" value in the table position defined by the entering variable (X_2) and the leaving variable (A_3). This is the so-called "pivot element". In addition, the table is modified by necessary algebraic row operations to create zero's in all of the other positions in the column corresponding to the entering variable. Referring to Table 16-3 we need to manipulate this linear system in such a way as to create a 0 in row (1) and row (0) of column X_2. To create the 0 in row (1), we multiply the "operator row" (3), which contains the pivot element, by -1 and add the respective values to those of row (1):

Table 16-4

ROW	BASIS	Z	X_1	X_2	A_1	S_2	E_3	A_3	b	r
(0)	Z	-1	3-M	1-2M	0	0	M	0	-1-4M	
(1)	A_1	0	1	0	1	0	1	-1	2	
(2)	S_2	0	1	0	0	1	0	0	5	
(3)	X_2	0	0	1	0	0	-1	1	1	

Similarly, to create a zero in row (0) of column X_2, the operator row (3) is multiplied by (-1+2M) and added to the corresponding elements of row (0):

Table 16-5

ROW	BASIS	Z	X_1	X_2	A_1	S_2	E_3	A_3	b	r
(0)	Z	-1	3-M	0	0	0	1-M	-1+2M	-2-2M	
(1)	A_1	0	1	0	1	0	1	-1	2	2
(2)	S_2	0	1	0	0	1	0	0	5	-
(3)	X_2	0	0	1	0	0	-1	1	1	-

One iteration of the simplex procedure has been completed. We now return to step (2) to determine if this new solution is optimal and continue the process until the optimal solution is determined. Inspecting row (0) in Table 16-5 it is apparent that the solution is not optimal because not all of the values in row (0) are ≥ 0. Since the most negative coefficient corresponds to E_3, it enters the basis. Since there is only one positive a_{ij}^* value in the E_3 column, corresponding to A_1, A_1 must be the leaving variable. To reflect this new basis, the system is manipulated to include E_3 in the identity

matrix. To create a zero in row (3), column E_3, the operator row (1) is added to row (3):

Table 16-6

ROW	BASIS	Z	X_1	X_2	A_1	S_2	E_3	A_3	b	r
(0)	Z	-1	3-M	0	0	0	1-M	-1+2M	-2-2M	
(1)	E_3	0	1	0	1	0	1	-1	2	
(2)	S_2	0	1	0	0	1	0	0	5	
(3)	X_2	0	1	1	1	0	0	0	3	

Then, to create a zero in row (0), column E_3, multiply the elements of the operator row (1) by (-1+M) and add the respective to those in row (0):

Table 16-7

ROW	BASIS	Z	X_1	X_2	A_1	S_2	E_3	A_3	b	r
(0)	Z	-1	2	0	-1+M	0	0	M	0	
(1)	E_3	0	1	0	1	0	1	-1	2	
(2)	S_2	0	1	0	0	1	0	0	5	
(3)	X_2	0	1	1	1	0	0	0	3	

Inspecting row (0), all coefficients for the nonbasic variables are > 0. This solution is optimal. The optimal policy then is to let $X_1 = 0$ and $X_2 = 3$. This results in a value of $Z = 0$. In addition, inspection of this final table indicates several other important pieces of information. It suggests that under this optimal policy, $E_3 = 2$ or that 2 "extra" units of the requirement stipulated by constraint (3) will be produced. In addition, since, under the optimal policy, $S_2 = 5$, 5 units of the resource represented by constraint (2) will remain unused. While the simplex procedure may appear to be rather cumbersome to accomplish manually, it is readily performed by the computer, and many very efficient LP computer codes are available.

Abberations in the Simplex Algorithm

In earlier discussion, a number of abberations were introduced, e.g., alternative optimal solutions, unbounded solutions, and no feasible solution. How are these conditions identified in the simplex procedure?

Alternative Optimal Solutions

Recall that in the two variable situation, alternative optima exist when the objective function is parallel to a binding constraint. Consider the example:

$$\text{maximize:} \quad Z = X_1 + 3X_2$$

$$\begin{aligned} \text{s.t.} \quad 2X_2 + 6X_2 &\le 12 \\ 3X_1 + 2X_2 &\le 6 \\ X_1,\ X2_2 &\ge 0 \end{aligned}$$

Converting to standard form:

$$\text{maximize:} \quad Z = X_1 + 3X_2 + OS_1 + OS_2$$

$$\begin{aligned} \text{s.t.} \quad & 2X_1 + 6X_2 + S_1 = 12 \\ & 3X_1 + 2X_2 + S_2 = 6 \\ & X_1, X_2, S_1, S_2 \geq 0 \end{aligned}$$

In tableau form, this becomes

Table 16-8

ROW	BASIS	Z	X_1	X_2	S_1	S_2	b_i	r_i
(0)	Z	1	-1	-3	0	0	0	
(1)	S_1	0	2	6	1	0	12	2
(2)	S_2	0	3	2	0	1	6	3

The initial basic feasible solution is not optimal. X_2 is the entering variable. As X_2 enters, S_1 is driven out of solution first ($r_i = 2$) so it (S_1) is the leaving variable. Adjusting the table, we need to create a "1" in the pivot element position of row (1), column X_2, and a "0" in the other two positions of this column. The new solution is:

Table 16-9

ROW	BASIS	Z	X_1	X_2	S_1	S_2	b_i	r_i
(0)	Z	1	0	0	1/2	0	6	
(1)	X_2	0	1/3	1	1/6	0	2	
(2)	S_2	0	7/3	0	-2/6	1	2	

Since all the coefficients in row (0) are ≥ 0, this solution is optimal. Now, note that nonbasic variable X_1 has a coefficient of 0 in row (0) of the optimal solution. Bringing X_1 into the solution has no effect on Z, implying that alternative optimal solutions exist.

Unbounded Solutions

Consider again the following example:

$$\text{maximize:} \quad Z = 2X_1 + X_1$$

$$\begin{aligned} \text{s.t.} \quad & X_1 + X_2 \geq 5 \\ & X_1 \geq 2 \\ & X_2 \geq 1 \end{aligned}$$

Converting to standard form:

$$\text{maximize:} \quad Z = 2X_1 + X_2 + OE_1 - MA_1 + OE_2 - MA_2 + OE_3 - MA_3$$

$$\begin{aligned} \text{s.t.} \quad & X_1 + X_2 - E_1 + A_1 = 5 \quad (1) \\ & X_1 - E_2 + A_2 = 2 \quad (2) \\ & X_2 - E_3 + A_3 = 1 \quad (3) \\ & X_j, E_i, A_i \geq 0 \quad (j = 1,2; \ i = 1,2,3) \end{aligned}$$

In table form:

Table 16-10

ROW	BASIS	Z	X_1	X_2	E_1	A_1	E_2	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	-2	-1	0	M	0	M	0	M	0	
(1)		0	1	1	-1	1	0	0	0	0	5	
(2)		0	1	0	0	0	-1	1	0	0	2	
(3)		0	0	1	0	0	0	0	-1	1	1	

To obtain an initial basic feasible solution, create a zero in row (0) for columns A_1, A_2, and A_3:

Table 16-11

ROW	BASIS	Z	X_1	X_2	E_1	A_1	E_2	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	-2-2M	-1-2M	M	0	M	0	M	0	-8M	
(1)	A_1	0	1	1	-1	1	0	0	0	0	5	5
(2)	A_2	0	1	0	0	0	1	1	0	0	2	2
(3)	A_3	0	0	1	0	0	0	0	-1	1	1	

The initial basic feasible solution is not optimal. X_1 enters and A_2 leaves. The new solution is:

Table 16-12

ROW	BASIS	Z	X_1	X_2	E_1	A_1	E_2	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	-1-2M	M	0	-2-M	-2+2M	M	0	4-4M	
(1)	A_1	0	0	1	-1	1	1	-1	0	0	3	3
(2)	X_1	0	1	0	0	0	-1	1	0	0	2	-
(3)	A_3	0	0	1	0	0	0	0	-1	1	1	1

This solution is not optimal. X_2 enters and A_3 leaves. The new solution is:

Table 16-13

ROW	BASIS	Z	X_1	X_2	E_1	A_1	E_2	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	0	M	0	-2-M	2+2M	-1-2M	1+2M	5-2M	
(1)	A_1	0	0	0	-1	1	1	-1	1	-1	2	2
(2)	X_1	0	1	0	0	0	-1	1	0	0	2	
(3)	X_2	0	0	1	0	0	0	0	-1	1	1	

The solution is not yet optimal. E_3 enters and A_1 leaves. The new solution is:

Table 16-14

ROW	BASIS	Z	X_1	X_2	E_1	A_1	E_2	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	0	-1-M	1+2M	-1+M	1	0	0	7+2M	
(1)	E_3	0	0	0	-1	1	1	-1	1	-1	2	
(2)	X_1	0	1	0	0	0	-1	1	0	0	2	
(3)	X_2	0	0	1	-1	1	1	-1	0	0	3	

The solution is still not optimal. E_1 is the entering variable. However, all of the a_{ij} values in the column vector for E_1 are ≤ 0. It is not possible to compute a positive r_i for any row. This indicates that bringing E_1 into the solution basis will not drive any of the existing basic variables out. The solution is unbounded (in the direction of maximization).

No Feasible Solution

Consider again the problem:

$$\text{maximize:} \quad Z = 2X_1 + X_2 \qquad (0)$$

$$\text{s.t.} \quad \begin{aligned} X_1 + X_2 &\leq 4 && (1) \\ X_1 \quad &\geq 5 && (2) \\ X_2 &\geq 6 && (3) \end{aligned}$$

Rewriting in standard form:

$$\text{maximize:} \quad Z = 2X_1 + X_2 + OS_1 + OE_2 - MA_2 + OE_3 - MA_3$$

$$\text{s.t.} \quad \begin{aligned} X_1 + X_2 + S_1 \qquad\qquad &= 4 \\ X_1 \qquad - E_2 + A_2 \quad &= 5 \\ X_2 \qquad - E_3 + A_3 &= 6 \end{aligned}$$

In tableau form:

Table 16-15

ROW	BASIS	Z	X_1	X_2	S_1	E_2	A_2	E_3	A_3	b_i	r_i
(0)		1	-2	-1	0	0	M	0	M	0	
(1)		0	1	1	1	0	0	0	0	4	
(2)		0	1	0	0	-1	1	0	0	5	
(3)		0	0	1	0	0	0	-1	1	6	

To obtain an initial basic feasible solution, create a zero in row (0) for columns A_2 and A_3:

Table 16-16

ROW	BASIS	Z	X_1	X_2	S_1	E_2	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	-2-M	-1-M	0	M	0	M	0	-11M	
(1)	S_1	0	1	1	1	0	0	0	0	4	4
(2)	A_2	0	1	0	0	-1	1	0	0	5	5
(3)	A_3	0	0	1	0	0	0	-1	1	6	-

The solution is not optimal. X_1 enters and S_2 leaves. The new solution is:

Table 16-17

ROW	BASIS	Z	X_1	X_2	S_1	E_2	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	1	2+M	M	0	M	0	8+7M	
(1)	X_1	0	1	1	1	0	0	0	0	4	
(2)	A_2	0	0	-1	-1	-1	1	0	0	1	
(3)	A_3	0	0	1	0	0	0	-1	1	6	

Looking at row (0), all entries are ≥ 0. The solution is optimal. However, the solution basis includes artificial variables having positive values. Whenever an artificial variable appears in the optimal solution basis at a strictly positive level, there is no feasible solution. If an artificial variable appears in the optimal solution basis and is equal to zero, the solution is feasible, although the model may have a redundant constraint or basic feasible solutions which are degenerate.

Degeneracy

As previously defined, a solution is described as degenerate when one or more of the basic variables (in addition to the n nonbasic variables arbitrarily set equal to zero) equal zero. In the simplex procedure, this condition is indicated when there is a tie for the leaving variable, i.e., as the entering variable value is increased, two (or more) basic variables are driven to zero and out of the basis simultaneously (two or more basic variables have the same r_i value). When this occurs, the tie is broken by arbitrarily selecting the leaving variable. (Note: If one of the variables is artificial and the other is not, you should select the artificial variable.) The only real problem with degeneracy is that in very rare situations, the simplex algorithm "cycles" and "stalls out," i.e., it returns to a previously generated nonoptimal solution rather than toward the optimal solution. This condition can be corrected, in some cases, by simply selecting another one of the tied variables. When this remedy does not correct the situation, more advanced procedures, not described here, are available.

POSTOPTIMALITY (SENSITIVITY) ANALYSIS

Introduction

In constructing an LP model for a decision system, we assume that all system relationships are linear and that the model parameters (c_j, b_i, and a_{ij}) are constant and known with certainty. Obviously, the solution of an LP problem can be no better than the quality (validity) of the assumptions upon which the model is based. It is inconsistent (and very risky!) to place a high degree of confidence in a decision policy which has been derived from a system model for which the parameters, components, and relationships are not known with equal confidence. Having obtained an optimal solution to an LP model, the analyst needs to investigate the "sensitivity" of this optimal policy to changes in assumed relationships and, in particular, to changes in values assumed for the respective model parameters. If the solution is relatively "insensitive" to variation in model relationships and parameters, then the analyst might feel reasonably confident in the optimal solution - knowing that variations in the model assumptions do not significantly affect the solution policy and criterion value. On the other hand, if postoptimality analysis reveals that the optimal solution is sensitive and varies significantly with relatively minor variations in the model assumptions, the decision taker should proceed with caution. When the solution is sensitive and the decision is important (in terms of mission performance effectiveness, cost, safety, etc.) it would probably be prudent to acquire, if possible, the additional information necessary to improve our knowledge of the actual value of system model parameters and system relationships. Of course, knowing when the model is "good enough" and, relatedly, how confident one needs to be in the quality of the model-derived decision policy before committing to action is, ultimately, a function of the decision taker's skill, experience, propensity for, or adversion to, risk, and value system.

This discussion focuses primarily on parameter sensitivity. However, the analyst should also (at least) consider the sensitivity of the optimal policy to adding or deleting system model components (variables) and/or relationships (constraints). Also, the sensitivity analysis techniques presented in this section are based on varying one parameter at a time. Certainly, the simultaneous variation of more than a single parameter is more realistic. However, the methodology for such "multivariate" sensitivity analysis is rather complex and beyond the scope of this discussion. For a more complete and detailed treatment of postoptimality (sensitivity), you should refer to one of the references following this section.

As the name suggests, postoptimality analysis begins after the optimal basic feasible solution to the decision system model has been determined, i.e., we begin with the solution represented by the final table or tableau. To illustrate this discussion, consider the following example:

maximize: $Z = 6X_1 + X_2$

$$\text{s.t.} \quad \begin{aligned} 5X_1 + 6X_2 &\le 30 \\ 2X_1 + X_2 &= 8 \\ X_1 + \quad &\ge 1 \\ X_1, X_2 &\ge 0 \end{aligned}$$

To solve this model, convert to standard form:

$$\text{maximize:} \quad Z = 6X_1 + X_2 + OS_1 - MA_2 + OE_3 - MA_3 \qquad (0)$$

$$\text{s.t.} \quad \begin{aligned} 5X_1 + 6X_2 + S_1 \quad &= 30 \qquad (1) \\ 2X_1 + X_2 + A_2 \quad &= 8 \qquad (2) \\ X_1 \quad - E_3 + A_3 &= 1 \qquad (3) \\ X_1, X_2, A_2, E_3, A_3 &\ge 0 \end{aligned}$$

In tabular form:

Table 16-18

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)		1	-6	-1	0	M	0	M	0	
(1)		0	5	6	1	0	0	0	30	
(2)		0	2	1	0	1	0	0	8	
(3)		0	1	0	0	0	-1	1	1	

To develop an initial basic feasible solution, eliminate the "M" values from row (0):

Table 16-19

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	-6-3M	-1-M	0	0	M	0	-9M	
(1)	S_1	0	5	6	1	0	0	0	30	6
(2)	A_2	0	2	1	0	1	0	0	8	4
(3)	A_3	0	1	0	0	0	-1	1	1	1

The solution is not optimal. X_1 enters and A_3 leaves. The new solution is:

Table 16-20

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	-1-M	0	0	-6-2M	6+3M	6-6M	
(1)	S_1	0	0	6	1	0	5	-5	25	5
(2)	A_2	0	0	1	0	1	2	-2	6	3
(3)	X_1	0	1	0	0	0	-1	1	1	-

The solution is not optimal. E_3 enters and A_2 leaves. The new solution is:

Table 16-21

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	2	0	3+M	0	M	24	
(1)	S_1	0	0	7/2	1	-5/2	0	0	10	
(2)	E_3	0	0	1/2	0	1/2	1	-1	3	
(3)	X_1	0	1	1/2	0	1/2	0	0	4	

The solution is optimal. This final tableau contains a wealth of information for the manager. By inspection, you can see that the optimal policy is $X_1 = 4$, and $X_2 = 0$, resulting in $Z = 24$. The final table also shows that, since $S_1 = 10$, only 20 units of the 30 units of resource (1) originally available were consumed. The slack of 10 means these resources are available for use by the manager elsewhere. Similarly, the final table also indicates $E_3 = 3$, i.e., in the optimal solution, 3 more units than are required by constraint (3) are produced. Additional information about the sensitivity of the optimal solution policy to variations in the assumed value of the respective model parameters (c_j, b_i, a_{ij}) can be obtained through analysis of the final (optimal) table.

Coefficients of the Objective Function (C_j)

In performing sensitivity analysis on the coefficients in the objective function (C_j), the general objective is to determine the range over which the respective variable coefficient can vary without changing the optimal solution. The specific analysis procedure used depends on whether we are considering the coefficient of a nonbasic or basic variable.

Nonbasic Variables

Recall that the simplex procedure terminates when the test for optimality indicates that all of the coefficients in row (0) are ≥ 0. From the table corresponding to the optimal solution in the example, you can see by inspection that X_1 is a basic variable while X_2 is nonbasic. Over what range (ΔC_2) would C_2 have to vary before X_2 becomes a basic variable, changing the optimal solution? In order for X_2 to enter, its corresponding coefficient would have to be <u>negative</u>, indicating the basis is not optimal. Introducing ΔC_2 into the table:

Table 16-22

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	$2-\Delta C_2$	0	3+M	0	M	24	
(1)	S_1	0	0	7/2	1	-5/2	0	0	10	
(2)	E_3	0	0	1/2	0	1/2	1	-1	3	
(3)	X_1	0	1	1/2	0	1/2	0	0	4	

We can write $2 - \Delta C_2 > 0$ or, alternatively, $\Delta C_2 < 2$. Since, in the original problem, $C_2 = 1$, then $C_2 + \Delta C_2 < 1+2$ or <3. This means that the current solution basis will remain optimal as long as $C_2 < 3$. Note that there is no lower limit to this range. <u>The range over which the objective function coefficient of a nonbasic decision variable can vary without changing the optimal solution is called the range of insignificance</u>.

Basic Variables

You perhaps noted that introducing a "Δ" term in the table for a nonbasic variable had no effect on the values of the basic variables or on the value of the criterion variable Z. Such is not the case when we attempt sensitivity

analysis on the objective function coefficient for a basic decision variable. Beginning in the same manner as we did with the nonbasic variable, a "Δ" is introduced in row (0) for the basic variable being considered (e.g., X_1):

Table 16-23

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	$0-\Delta C_1$	2	0	3+M	0	M	24	
(1)	S_1	0	0	7/2	1	-5/2	0	0	10	
(2)	E_3	0	0	1/2	0	1/2	1	-1	3	
(3)	X_1	0	1	1/2	0	1/2	0	0	4	

Notice, however, that by introducing the ΔC_1 term, we no longer have the required identity matrix. The matrix needs to be manipulated to reestablish the identity condition. The revised table follows:

Table 16-24

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	$2+1/2\Delta C_1$	0	$3+M+1/2\Delta C_1$	0	M	$24+4\Delta C_1$	
(1)	S_1	0	0	7/2	1	5/2	0	0	10	
(2)	E_3	0	0	1/2	0	1/2	1	-1	3	
(3)	X_1	0	1	1/2	0	1/2	0	0	4	

In the process of eliminating the ΔC_1 term from the X_1 column, it was introduced in the X_2, A_2, and b_i columns. As in the case of the nonbasic variable, we can now write the conditions for preserving the current solution basis:

For X_2:

$$\begin{aligned} 2+1/2\Delta C_1 &> 0 \\ 1/2\Delta C_1 &> -2 \\ \Delta C_1 &> -4 \end{aligned}$$

For A_2:

$$\begin{aligned} 3+M+1/2\Delta C_1 &> 0 \\ 1/2\Delta C_1 &> -3-M \\ \Delta C_1 &> -6-2M \end{aligned}$$

Since the condition that $\Delta C_1 > -4$ also insures compliance with the second condition, i.e., $\Delta C_1 > -6-2M$, this second constraint is redundant. Consequently, since $C_1 = 6$ in the model:

$$C_1 + \Delta C_1 > 6 - 4 \quad \text{or} \quad C_1 + \Delta C_1 > 2$$

Therefore, as long as the value of C_1 remains > 2, the optimal solution basis will not change. However, because we are dealing with a basic variable and Z is now equal to $24 + 4\Delta C_1$, varying the value of C_1 will be reflected in the value of Z. Since $\Delta C_1 > -4$, Z could vary as low as $24 + 4(-4) = 24 - 16 = 8$, without reflecting a change in the optimal solution basis. <u>The range over which the objective function coefficient of a basic variable can vary without changing the optimal solution basis is called the range of optimality.</u>

Boundary Value Constants (b_i)

In conducting a sensitivity analysis on the boundary value or right-hand-side (b_i) constants, the objective is to determine the range over which the b_i constant can vary without changing the optimal solution basis.

Let's consider the first constraint: $5X_1 + 6X_2 \leq 30$. We would like to know the range on b such that the optimal solution remains unchanged. For purposes of illustration, let's rework the example problem, replacing b with $b_1 + \Delta b_1$:

Table 16-25

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)		1	-6	-1	0	M	0	M	0	
(1)		0	5	6	1	0	0	0	$30+\Delta b_i$	
(2)		0	2	1	0	1	0	0	8	
(3)		0	1	0	0	0	-1	1	1	

Clearing out the "M" values from row (0) to create the identity matrix:

Table 16-26

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	-6-3M	-1-M	0	0	M	0	-9M	
(1)	S_1	0	5	6	1	0	0	0	$30+\Delta b_i$	$6+\frac{\Delta b_i}{5}$
(2)	A_2	0	2	1	0	1	0	0	8	4
(3)	A_3	0	1	0	0	0	-1	1	1	1

The solution is not optimal. X_1 enters, A_3 leaves. The new solution is:

Table 16-27

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	-1-M	0	0	-6-2M	6+3M	6-6M	
(1)	S_1	0	0	6	1	0	5	-5	$25+\Delta b_i$	$5+\frac{\Delta b_i}{5}$
(2)	A_2	0	0	1	0	1	2	-2	6	3
(3)	X_1	0	1	0	0	0	-1	1	1	-

The solution is not optimal. E_3 enters and A_2 leaves. The new solution is:

Table 16-28

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i	r_i
(0)	Z	1	0	2	0	3+M	0	M	24	
(1)	S_1	0	0	7/2	1	-5/2	0	0	$10+\Delta b_i$	
(2)	E_3	0	0	1/2	0	1/2	1	-1	3	
(3)	X_1	0	1	1/2	0	1/2	0	0	4	

The solution is optimal. It is not necessary to completely rework the problem as was just done for purposes of illustration. Since, in the process of solving this problem, the same algebraic row operations were performed on both the left-hand and right-hand sides of the equation, the Δb_1 terms in column b_1 will have exactly the same coefficients as the vector S_1 (S_1 being the slack associated with constraint (1) and, therefore, b_1):

Table 16-29

ROW	BASIS	Z	X_1	X_2	S_1	A_2	E_3	A_3	b_i
(0)	Z	1	0	2	0	3+M	0	M	$24+0\Delta b_1$
(1)	S_1	0	0	7/2	1	-5/2	0	0	$10+1\Delta b_1$
(2)	E_3	0	0	1/2	0	1/2	1	-1	$3+0\Delta b_1$
(3)	X_1	0	1	1/2	0	1/2	0	0	$4+0\Delta b_1$

Now, recalling that the basic variables will remain in the solution basis as long as they are >0, we can write: $10 + \Delta b_1 > 0$ or $\Delta b_1 > -10$ since $b_1 = 30$ in the original problem, as long as $b_1 > (b_1 + \Delta b_1) > (30 - 10) > 20$, the solution basis will remain unchanged. Now, consider the second constraint: $2X_1 + X_2 = 8$. In this case, we introduced the <u>artificial</u> variable A_2 to obtain an initial identity matrix. Consequently, the coefficients appearing under A_2 in the optimal solution are the ones to use as coefficients for the Δb_2 term under the b_i column. Recognizing this, you can write:

(0) $24 + (3 + M)\Delta b_2$

(1) $10 + (-5/2)\Delta b_2 > 0$

(2) $3 + 1/2\ \Delta b_2 > 0$

(3) $4 + 1/2\ \Delta b_2 > 0$

From (1): $\Delta b_2 < 4$
From (2): $\Delta b_2 > -6$
From (3): $\Delta b_2 > -8$

Combining these: $-6 < \Delta b_2 < 4$

Now, since $b_1 = 8$: $2 < b_2 < 12$

Note that as long as b_2 remains in the range $2 < b_2 < 12$ (or alternatively, $-6 < \Delta b_2 < 4$) the optimal solution basis remains unchanged. However, the <u>value</u> of these basic variables can change as b_2 varies. Suppose, for example, b_2 is changed to 10 ($\Delta b_2 = 2$). We can assess the effect of this change without resolving the problem:

(0) $Z = 24 + 3(2) = 30$ *

(1) $S = 10 - 5/2(2) = 5$

(2) $E = 3 + 1/2(2) = 4$

(3) $X = 4 + 1/2(2) = 5$

*Note: the "M" term in the coefficient (3 + M) is ignored.

If, however, we want to assess the effect of $\Delta b_2 = 6$, this exceeds the previously determined range and the problem must be resolved. Finally, consider the third constraint: $X_1 \geq 1$. Here, too, an artificial variable (A_3) was used in the initial solution basis. The coefficients under A_3 in the final tableau are the coefficients to use for the coefficients of Δb_3 in column b_i:

(0) $24 + M\Delta b_3$

(1) $10 + 0\Delta b > 0$

(2) $3 - 1\Delta b > 0$

(3) $4 + 0\Delta b > 0$

From (2): $\Delta b_3 < 3$

Therefore, since $b_3 = 1$, $b_3 < 4$.

The range over which a right-hand constant (b_i) can vary without changing the optimal solution basis is called the range of feasibility.

Shadow Prices

The concept of the range of feasibility for b_i suggests a very important related notion - the "shadow price." The solution to every LP model is determined by the intersection of two or more limiting or binding constraints, i.e., constraints for which all of the available resource (b_i) is consumed in the optimal solution or for which demand is satisfied without surplus. In this case, the ability to improve Z even further is constrained. As a manager, it is of interest to know how much it is worth (in terms of Z) to have an additional unit of the constraining resource or requirement. This value is called the shadow price for the ith constraint. The shadow price can be read directly from the final (optimal) solution table. The shadow price associated with the ith constraint is the coefficient in row (0) underneath the associated slack variable for $\leq$ constraints and the pure number (ignoring "M" terms) for the artificial variable associated with (=) or ($\geq$) constraints. The shadow price holds throughout the range of feasibility for that constraint. Consider, for example, the final tableau:

Table 16-30

ROW	BASIS	Z	X_1	X_2	S_1	S_2	A_3	b_i	r_i
(0)	Z	1	0	0	2	0	3+M	38	
(1)	X_1	0	1	0	1	0	-1	4	
(2)	S_2	0	0	0	1	1	-3	2	
(3)	X_2	0	0	1	-1	0	2	2	

The shadow price for constraint (1) = 2. The range of feasibility for this constraint and shadow price is:

(0) $38 + 2\Delta b_1$

(1) $4 + 1\Delta b_1 > 0 \qquad \Delta b_1 > -4$

(2) $2 + 1\Delta b_1 > 0 \qquad \Delta b_1 > -2$

(3) $2 + 1\Delta b_1 > 0 \qquad \Delta b_1 < 2$

$$-2 < \Delta b < 2$$

For constraint (2), the shadow price is 0 (as you would expect because $S_2 = 2$, i.e., we currently have 2 excess units of commodity (2), so why buy any more?). The range of feasibility for (2) is:

(0) $38 + 0\Delta b_2$

(1) $4 + 0\Delta b_2 > 0$

(2) $2 + 1\Delta b_2 > 0 \qquad \Delta b_2 > -2$

(3) $2 + 0\Delta b_2 > 0$

For constraint (3), the shadow price is 3, corresponding to the pure number portion of the quantity (3 + M). The corresponding range of feasibility is:

(0) $38 + 3\Delta b_3$

(1) $4 - 1\Delta b_3 > 0 \qquad \Delta b_3 < 4$

(2) $2 - 3\Delta b_3 > 0 \qquad \Delta b_3 < 2/3$

(3) $2 + 2\Delta b_3 > 0 \qquad \Delta b_3 > -1$

$$-1 < \Delta b_3 < 2/3$$

Other Postoptimality Analyses

In addition to examining the ranges of insignificance, optimality, and feasibility, there are other types of postoptimality analysis that can be accomplished. For example, sensitivity analysis can be accomplished on the constraint coefficients (a_{ij}). This analysis effectively seeks to determine the sensitivity of the optimal solution to variation in the slope of the respective constraints. Computationally, this analysis is somewhat more difficult, the difficulty depending on whether basic or nonbasic variables are being considered. As suggested previously, it is also possible to analytically assess the sensitivity of the optimal solution to the simultaneous variation of two or more parameters, as well as to examine the sensitivity of the solution to the addition or deletion of constraints in the decision system model.

To summarize and reiterate, mathematical modeling, in general, and LP, in particular, is a process which includes a systematic analysis of the decision situation, constructing a mathematical model that adequately captures the essence of the situation, and solving the model. Equally important, the process also includes interpreting the model solution back to its parent reality, in part by conducting postoptimality analysis to assess the sensitivity of the model-derived optimal policy to changes in model parameters. It is important to recognize that a manager's decision-taking effectiveness is affected to a large extent by the model used to assess the situation and to derive a particular course of action, whether that model is an intuitive mental image or a more systematic and rigorous formulation.

INTEGRATIVE EXAMPLE

The following is provided to illustrate the total modeling process, i.e., system analysis, model formulation, model solution, and postoptimality sensitivity analysis.

Problem Statement

A particular aircraft system corrosion control solution is made from three basic concentrated ingredients: X_1, X_2, and X_3. A batch of the solution must include at least 20 ounces of the concentrates, but not more than 40 ounces. Further, for each ounce of X_1, at least 1/4 ounce of X_3 must be used. Similarly, at least 1/2 ounce of X_2 should be used for each ounce of X_1. The costs of the concentrates X_1, X_2, and X_3 are \$5, \$15, and \$20 per ounce, respectively.

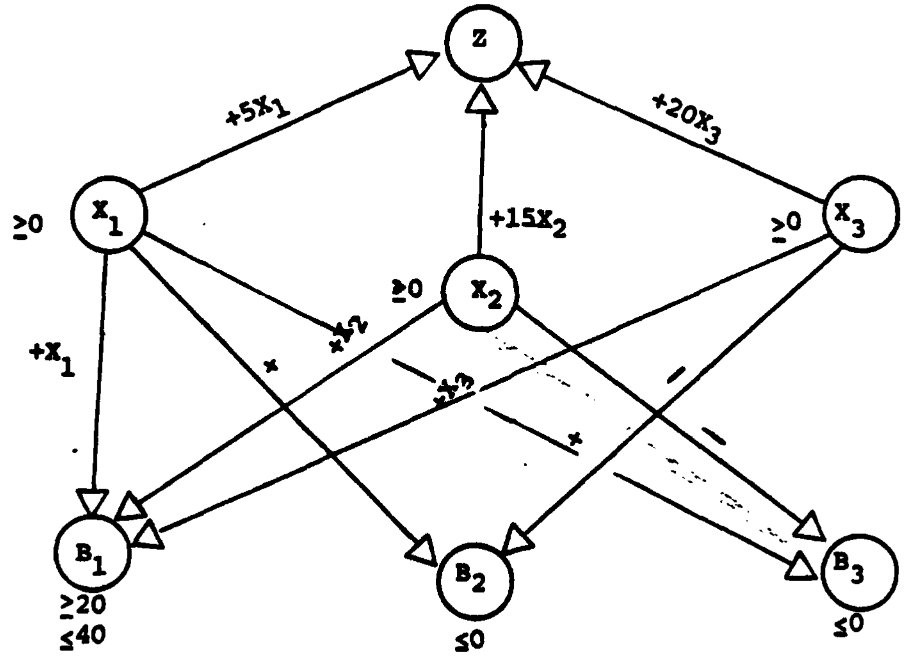
System Analysis

System Components:

1. Criterion Variable (Z): Total cost for a batch of solution in dollars; the objective is to minimize Z.
2. Decision Variables (X_j): X_j represents the ounces of ingredient j used in solution (j = 1,2,3)
3. Environmental Variables (B_i)
   B_1: total ounces of concentrates ($20 \leq B_1 \leq 40$)
   B_2: ratio of X_1 to X_3 ($B_2 \leq 4$)
   B_3: ratio of X_1 to X_2 ($B_3 \leq 2$)

System Relationships

Figure 16-16



LP Model

minimize: $Z = 5X_1 + 15X_2 + 20X_3$

s.t.

$$\begin{aligned} X_1 + X_2 + X_3 &\geq 20 \\ X_1 + X_2 + X_3 &\leq 40 \\ X_1 - 4X_3 &\leq 0 \\ X_1 - 2X_2 &\leq 0 \\ X_1, X_2, X_3 &\geq 0 \end{aligned}$$

LP Model Solution

Standard Form:

$$\text{maximize} \quad (-Z) = -5X_1 - 15X_2 - 20X_3 - 0E_1 - MA_1 - 0S_2 - 0S_3 - 0S_4$$

$$\begin{aligned} \text{s.t.} \quad X_1 + X_2 + X_3 - E_1 + A_1 &= 20 \quad (1) \\ X_1 + X_2 + X_3 + S_2 &= 40 \quad (2) \\ X_1 - 4X_3 + S_3 &= 0 \quad (3) \\ X_1 - 2X_2 + S_4 &= 0 \quad (4) \\ X_1, X_2, X_3, E_1, A_1, S_1, S_2, S_3, S_4 &\geq 0 \end{aligned}$$

Convert to tableau form:

Table 16-31

ROW	BASIS	Z	X_1	X_2	X_3	E_1	A_1	S_2	S_3	S_4	b_i	r_i
(0)	Z	-1	+5	+15	+20	0	M	0	0	0	0	
(1)		0	1	1	1	-1	1	0	0	0	20	
(2)		0	1	1	1	0	0	1	0	0	40	
(3)		0	1	0	-4	0	0	0	1	0	0	
(4)		0	1	-2	0	0	0	0	0	1	0	

To create an initial basic feasible solution, eliminate M from row (0), column A_1:

Table 16-32

ROW	BASIS	Z	X_1	X_2	X_3	E_1	A_1	S_2	S_3	S_4	b_i	r_i
(0)	Z	-1	5-M	15-M	20-M	M	0	0	0	0	-20M	
(1)	A_1	0	1	1	1	-1	1	0	0	0	20	20
(2)	S_2	0	1	1	1	0	0	1	0	0	40	20
(3)	S_3	0	1	0	-4	0	0	0	1	0	0	0
(4)	S_4	0	1	-2	0	0	0	0	0	1	0	0

Solution is not optimal
X_1 enters
Tie for the leaving variable between S_3 and S_4; arbitrarily select S_3

Table 16-33

ROW	BASIS	Z	X_1	X_2	X_3	E_1	A_1	S_2	S_3	S_4	b_i	r_i
(0)	Z	-1	0	15-M	40-5M	M	0	0	-5+M	0	-20M	
(1)	A_1	0	0	1	5	-1	1	0	-1	0	20	4
(2)	S_2	0	0	1	5	0	0	1	0	0	40	8
(3)	X_1	0	1	0	-4	0	0	0	1	0	0	-
(4)	S_4	0	0	-2	4	0	0	0	1	1	0	0

Solution is not optimal
X_3 enters
S_4 leaves

Table 16-34

ROW	BASIS	Z	X_1	X_2	X_3	E_1	A_1	S_2	S_3	S_4	b_i	r_i
				35					-15	-10		
(0)	Z	-1	0	-7/2M	0	M	0	0	+9/4M	+5/4M	-20M	
(1)	A_1	0	0	7/2	0	-1	1	0	-9/4	-5/4	20	40/7
(2)	S_2	0	0	7/2	0	0	0	1	-5/4	-5/4	40	80/7
(3)	X_1	0	1	-2	0	0	0	0	2	1	0	
(4)	X_3	0	0	-1/2	1	0	0	0	1/4	1/4	0	

Solution is not optimal
X_2 enters
A_1 leaves

Table 16-35

ROW	BASIS	Z	X_1	X_2	X_3	E_1	A_1	S_2	S_3	S_4	b_i	r_i
(0)	Z	-1	0	0	0	10	-10+M	0	15/2	5/2	-200	
(1)	X_2	0	0	1	0	-2/7	2/7	0	-9/14	-5/14	40/7	
(2)	S_2	0	0	0	0	1	-1	1	1	0	20	
(3)	X_1	0	1	0	0	-4/7	4/7	0	5/7	2/7	80/7	
(4)	X_3	0	0	0	1	-1/7	1/7	0	-1/14	1/14	20/7	

Solution is optimal
$X_1 = 80/7$ $Z = 200$
$X_2 = 40/7$
$X_3 = 20/7$

__Postoptimality Sensitivity Analysis__

__Coefficients of the Objective Function__:

__Nonbasic Variables__: (none)

Basic Variables (X_1, X_2, X_3):

$C_1 = 5$	$10 - 4/7\Delta C_1 > 0$	$\Delta C_1 < 35/2$
	$-10 + M + 4/7\Delta C_1 > 0$	$\Delta C_1 > 7/4(-M+10)$
	$15/2 + 5/7 \ \Delta C_1 > 0$	$\Delta C_1 > -21/2$
	$5/2 + 2/7 \ \Delta C_1 > 0$	$\Delta C_1 > -35/4$
	$-35/4 < \Delta C_1 < 35/2$	
	$-15/4 < C_1 < 90/4$	

$C_2 = 15$	$10 - 2/7\Delta C_2 > 0$	$\Delta C_2 < 35$
	$-10 + M + 2/7\Delta C_2 > 0$	$\Delta C_2 > 7/2(10-M)$
	$15/2 - 9/14 \ \Delta C_2 > 0$	$\Delta C_2 < 35/3$
	$5/2 - 5/14 \ \Delta C_2 > 0$	$\Delta C_2 < 7/3$
	$\Delta C_2 < 7/3$	
	$C_1 < 17 \ 1/3$	

$C_3 = 20$	$10 - 1/7\Delta C_3 > 0$	$\Delta C_3 < 70$
	$-10 + M + 1/7\Delta C_3 > 0$	$\Delta C_3 > 7(10-M)$
	$15/2 - 1/14 \ \Delta C_3 > 0$	$\Delta C_3 < 105$
	$5/2 + 1/14 \ \Delta C_3 > 0$	$\Delta C_3 > -35$
	$-35 < \Delta C_3 < 70$	
	$-15 < C_3 < 90$	

Right-Hand-Side Constants (b_i)

$b_1 = 20$	$40/7 + 2/7\Delta b_1 > 0$	$\Delta b_1 > -20$
	$20 - \Delta b_1 > 0$	$\Delta b_1 < 20$
	$80/7 + 4/7\Delta b_1 > 0$	$\Delta b_1 > -20$
	$20/7 + 1/7\Delta b_1 > 0$	$\Delta b_1 > -20$
	$-20 < \Delta b_1 < 20$	
	$0 < b_1 < 40$	

$b_2 = 40$	$40/7 + 0 \ \Delta b_2 > 0$	
	$20 + 1 \ \Delta b_2 > 0$	$\Delta b_2 > -20$
	$80/7 + 0 \ \Delta b_2 > 0$	
	$20/7 + 0 \ \Delta b_2 > 0$	
	$\Delta b_2 > -20$	
	$b_2 > 20$	

$b_3 = 0$	$40/7 - 9/14\Delta b_3 > 0$	$\Delta b_3 < 80/9$
	$20 + 1\Delta b_3 > 0$	$\Delta b_3 > -20$
	$80/7 + 5/7 \ \Delta b_3 > 0$	$\Delta b_3 > -16$
	$20/7 - 1/14\Delta b_3 > 0$	$\Delta b_3 < 40$
	$-16 < \Delta b_3 < 80/9$	
	$-16 < b_3 < 80/9$	

$b_4 = 0$	$40/7 - 5/14\Delta b_4 > 0$	$\Delta b_4 < 16$
	$20 + 0\Delta b_4 > 0$	
	$80/7 + 2/7\Delta b_4 > 0$	$\Delta b_4 > -40$
	$20/7 + 1/14\Delta b_4 > 0$	$\Delta b_4 > -40$
	$-40 < \Delta b_4 < 16$	
	$-40 < b < 16$	

SPECIALIZED LP ALGORITHMS

Linear programming has achieved its popularity in part because of its wide applicability. Certain classes of LP problems have specialized structures which permit solution by algorithms which are computationally more efficient than the more generally applicable simplex method. The transportation problem and the assignment problem are included in this section because of their utility and applicability in commonly-encountered managerial decision-taking situations.

The Transportation Model

The transportation model refers to that class of decision situations in which some commodity is available in specified limited quantities at a number of supply points (origins) and is required in other specified amounts at a number of demand points (destinations). The model assumes that there is no difference in the quality of the commodity available at each supply point. The objective is to allocate the available supply to meet all demand requirements in such a manner as to, typically, minimize the total transportation cost or to maximize total profit associated with the allocation.

As with many other management science models, the transportation model takes its name from the prototype problem for which the model was developed. However, the structure of the classical transportation problem applies to many decision situations having nothing to do with the transportation or shipment of commodities. Because of this flexibility, the transportation model and, relatedly, the assignment model are often described by the more generic label: distribution models or systems.

General System Analysis

System Components:

1. Criterion Variable (Z): Total cost to distribute a particular commodity from supply points i ($i = 1,2,...,m$) to demand points j ($j=1,2,...,n$).

2. Decision Variable (X_{ij}): X_{ij} represents the units of commodity distributed from source i to demand point j, where $i = 1,2,...,m$ and $j = 1,2,...,n$

3. Environmental Variables:

S_i = amount of commodity allocated from origin point ($S_i = s_i$)

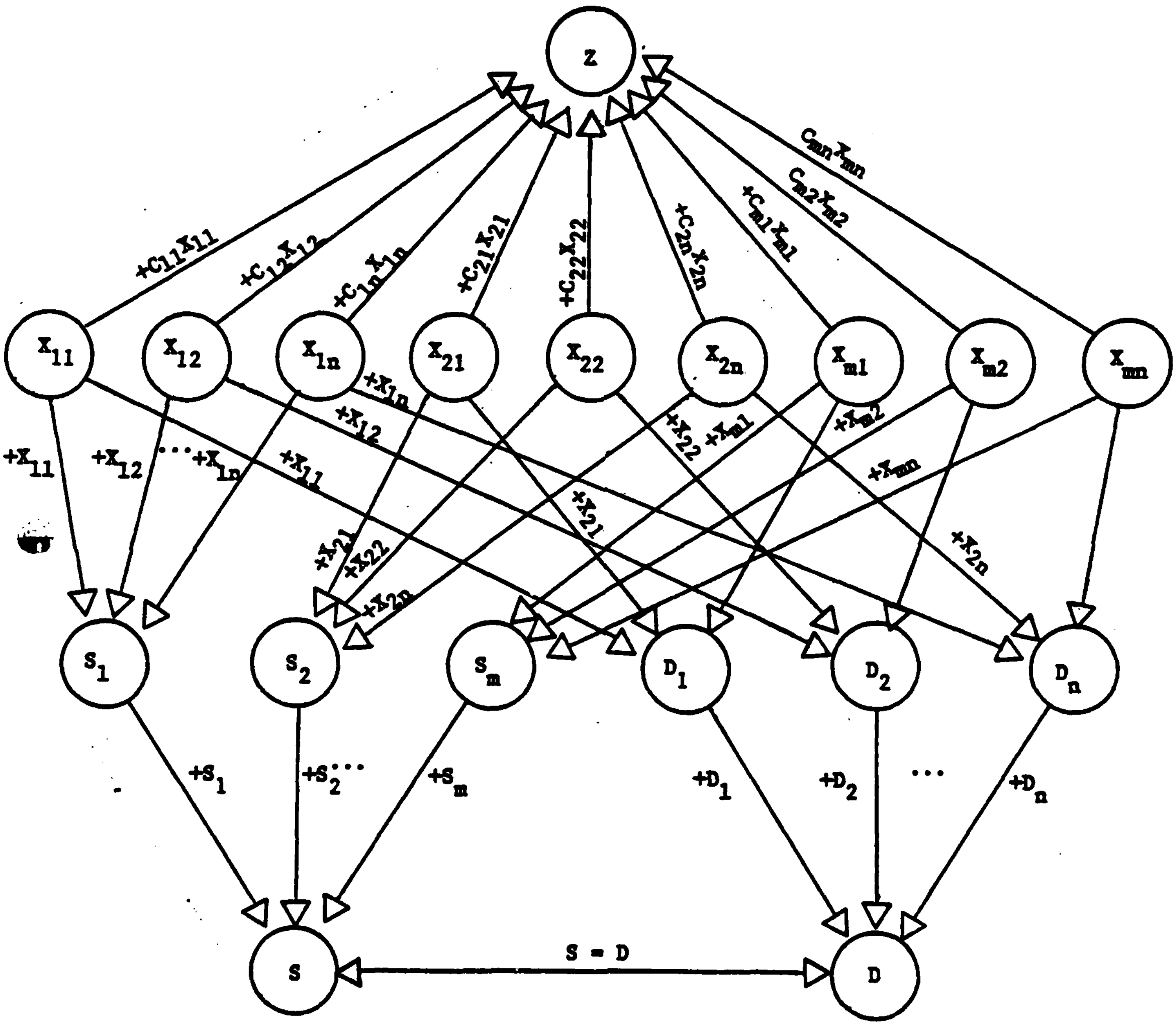
D_j = amount of commodity allocated to destination point ($D_j = d_j$)

S = total supply available at all sources

D = total demand required at all destinations

System Relationships

Figure 16-17



Mathematical Formulation

$$\text{minimize:} \quad Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} X_{ij} = S_i \; (i = 1,2,\ldots,m)$$

$$\sum_{i=1}^{m} X_{ij} = D_j \; (j = 1,2,\ldots n)$$

$$\sum_{i=1}^{m} S_i = \sum_{j=1}^{n} D_j$$

$$X_{ij} \geq 0$$

Note: Refer to page 16-58 for procedures to be followed in the more general case when total supply does not equal total demand, i.e., when

$$\sum_{i=1}^{m} S_i \neq \sum_{j=1} D_j$$

Formulation Examples

Example 16-4, is an example of a "classical" transportation problem involving the physical distribution of a commodity from supply points to demand points.

EXAMPLE 16-5

The Base Civil Engineer at Grand Forks AFB is planning the exterior painting work that must be done during the short summer period when this type of work can be satisfactorily accomplished. The planners, together with the foreman of the painting shop, have estimated the following requirements:

Base Facilities	7,800 manhours
Missile Sites	3,200 "
Family Housing	9,600 "

Labor is available from three sources:

In-House:	13,000 manhours	\$8/manhour
Overhires:	4,160 "	\$6/manhour
Contract:	3,440 "	\$10/manhour

Because of economic and security reasons, it is not practical to use contract labor on the missile site work. What manpower policy would you recommend?

System Analysis

System Components:

1. Criterion Variable (Z): Total labor cost in dollars to accomplish exterior painting requirements; the objective is to minimize total cost.

2. Decision Variable (X_{ij}): Manhours from source i used on task j, at a unit cost of C_{ij}, where:

i=		C_{ij}	j	
1.	In-house	8	1:	Base
2.	Overhire	6	2:	Missile
3.	Contract	10	3:	Housing

3. Environmental Variables:

S_1 = manhours of in-house labor used (S_1 = 13,000)

S_2 = manhours of overhire labor used (S_2 = 4,160)

S_3 = manhours of contracted labor used (S_3 = 3,440)

S = total manhours used from all sources (S = 20,600)

D_1 = total manhours used for base facilities (D_1 = 7,800)

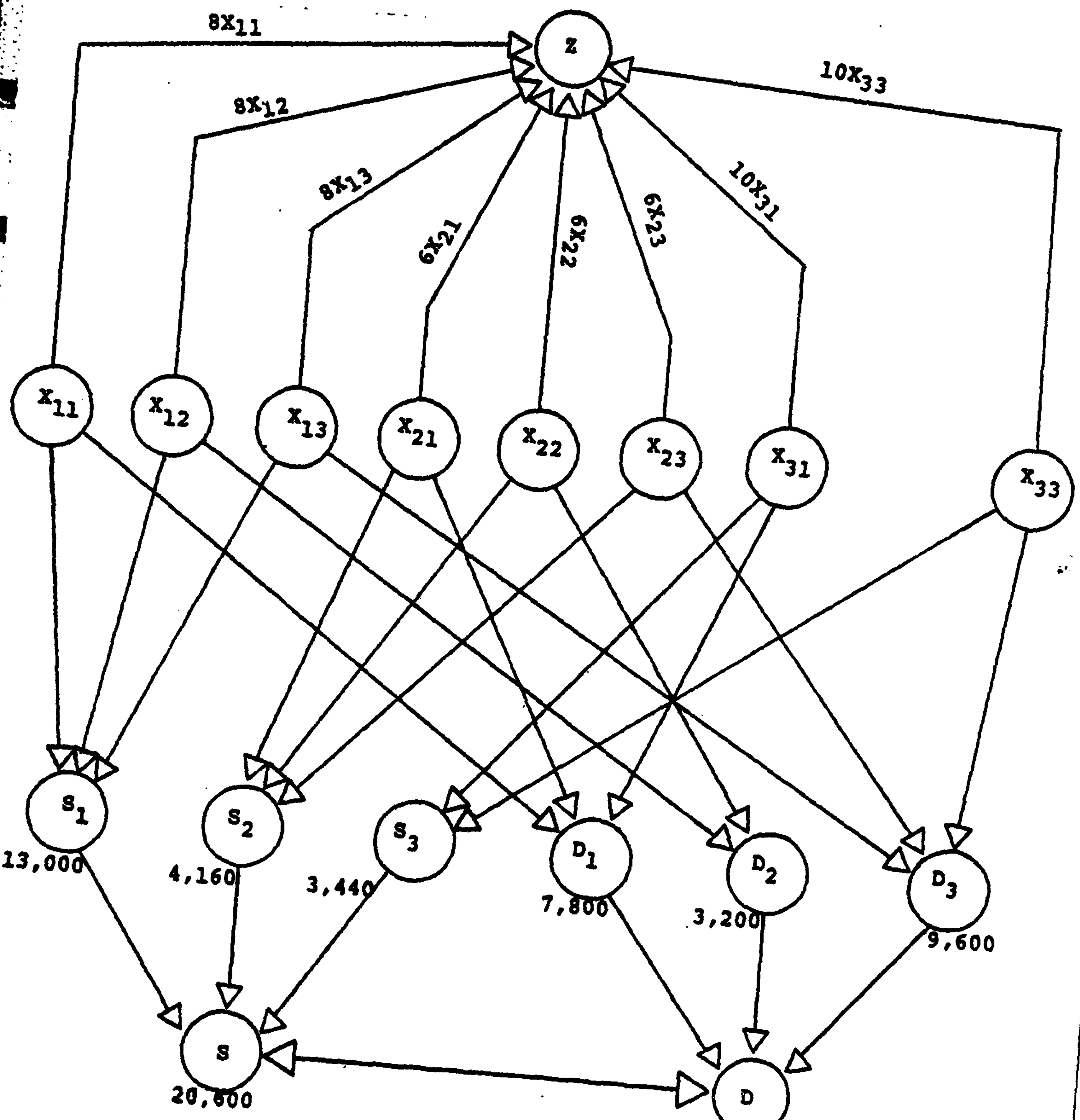
D_2 = total manhours used for missile sites (D_2 = 3,200)

D_3 = total manhours used for family housing (D_3 = 9,600)

D = total manhours used for all requirements (D = 20,600)

System Relationships:

Figure 16-18



Mathematical (LP) Model

$$\text{minimize:} \quad Z = 8X_{11} + 8X_{12} + 8X_{13} + 6X_{21} + 6X_{22} + 6X_{23} + 10X_{31} + 10X_{33}$$

$$\begin{aligned}
\text{s.t.} \quad & X_{11} + X_{12} + X_{13} = 13{,}000 \\
& X_{21} + X_{22} + X_{23} = 4{,}160 \\
& X_{31} + X_{33} = 9{,}600 \\
& X_{11} + X_{21} + X_{31} = 7{,}800 \\
& X_{12} + X_{22} = 3{,}200 \\
& X_{13} + X_{23} + X_{33} = 9{,}600 \\
& X_{ij} \geq 0 \quad (\text{all } i \text{ and } j)
\end{aligned}$$

EXAMPLE 16-6

You are the chief of a small engineering section primarily responsible for research and development on avionics systems. Currently, you have a staff of five design engineers whose experience and technical expertise vary. During the next three months, your office must complete five independent projects, each of which will consume the efforts of one of your staff engineers. In attempting to determine how the projects should be assigned, you have estimated, on a scale from one to ten, the efficiency (utility) of assigning a particular engineer to a specific project. These estimates are based on your assessment of the nature of each project (i.e., expertise required) and the skills and experience of your staff. The following table summarizes these estimates:

Table 16-36

		PROJECT				
		1	2	3	4	5
ENGINEER	1	6	8	9	6	2
	2	8	6	2	3	3
	3	6	5	8	7	6
	4	8	5	6	3	1
	5	8	6	5	6	0

How should the engineer-project assignments be made to maximize total utility?

System Analysis

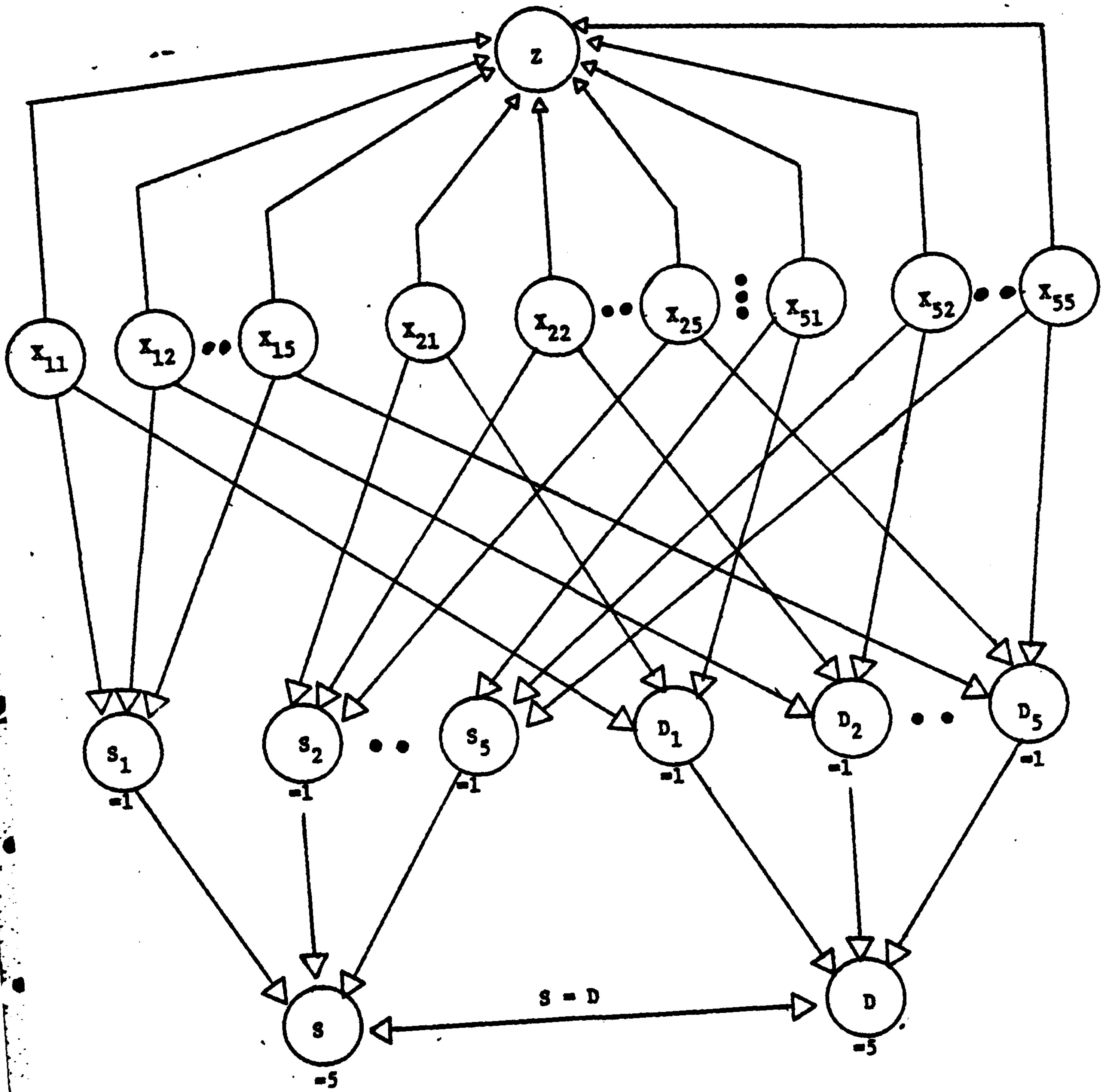
System Components:

1. Criterion Variable (Z): Total utility, maximize Z.
2. Decision Variable (X_{ij}): $X_{ij} = 1$ if engineer i is assigned to project j ($i, j = 1,2,\ldots,5$); $X_{ij} = 0$ if engineer i is not assigned to project j.
3. Environmental Variables:

S_i = no. of projects assigned to engineer i ($S_i = 1$), $i = 1,2,\ldots,5$
D_j = no. of engineers assigned to project j ($D_j = 1$), $j = 1,2,\ldots,5$
S = Total no. of engineers
D = Total no. of projects

System Relationships

Figure 16-19



Mathematical (LP) Model

$$
\begin{aligned}
\text{maximize:}\quad Z = {} & 6X_{11} + 8X_{12} + 9X_{13} + 6X_{14} + 2X_{15} \\
& + 8X_{21} + 6X_{22} + 2X_{23} + 3X_{24} + 3X_{25} \\
& + 6X_{31} + 5X_{32} + 8X_{33} + 7X_{34} + 6X_{35} \\
& + 8X_{41} + 5X_{42} + 6X_{43} + 3X_{44} + 1X_{45} \\
& + 8X_{51} + 6X_{52} + 5X_{53} + 6X_{54} + 0X_{55}
\end{aligned}
$$

$$
\begin{aligned}
\text{s.t.}\quad & X_{11} + X_{12} + X_{13} + X_{14} + X_{15} = 1 \\
& X_{21} + X_{22} + X_{23} + X_{24} + X_{25} = 1 \\
& \qquad\vdots \\
& X_{51} + X_{52} + X_{53} + X_{54} + X_{55} = 1 \\
& X_{11} + X_{21} + X_{31} + X_{41} + X_{51} = 1 \\
& X_{12} + X_{22} + X_{32} + X_{42} + X_{52} = 1 \\
& \qquad\vdots \\
& X_{15} + X_{25} + X_{35} + X_{45} + X_{55} = 1 \\
& X_{ij} = 0,1 \quad (i,\ j = 1,2,\ldots,5)
\end{aligned}
$$

Solution Techniques

Because the transportation model is a special case of the more general LP model, the simplex algorithm can be used to solve the transportation problem. In fact, conceptually, the steps in the transportation method (alternatively called the transportation simplex) are essentially the same as those in the basic simplex algorithm. Only the mechanics by which solution bases are created, examined for optimality, and revised are different.

The transportation method has two structural requirements which must be observed. First, total supply must equal total demand (effectively, this is the same as converting inequalities to equalities by adding slack or surplus variables). Secondly, the solution basis will never contain more than (m + n - 1) basic variables, where m equals the number of supply points or origins and n equals the number of demand points or destinations. This is due to the inherent redundancy in constraints created by the requirement that total supply equal total demand. When there are fewer than (m + n - 1) basic variables in solution (not equal to zero), the solution is degenerate.

The first step in the solution procedure is to convert the problem to a standard tabular form in which supply equals demand and the objective is one of minimization. The generalized tabular form for the transportation model is illustrated in Table 16-37.

Table 16-37

ORIGIN	DESTINATION				SUPPLY
	1	2	. . .	n	
1	C_{11} X_{11}	C_{12} X_{12}	. . .	C_{1n} X_{1n}	S_1
2	C_{21} X_{21}	C_{22} X_{22}	. . .	C_{2n} X_{2n}	S_2
⋮	⋮	⋮	⋮	⋮	⋮
m	C_{m1} X_{m1}	C_{m2} X_{m2}	. . .	C_{mn} X_{mn}	S_m
DEMAND	d_1	d_2	. . .	d_n	S D

To illustrate the procedure, consider the following mathematical (LP) model:

minimize: $Z = 4X_{11} + 8X_{12} + 8X_{13} + 16X_{21} + 24X_{22} + 16X_{23} + 8X_{31} + 16X_{32} + 24X_{33}$

s.t. $X_{11} + X_{12} \leq 76$

supply constraints:

$$
\begin{aligned}
X_{21} + X_{22} + X_{23} &\leq 82 \\
X_{31} + X_{32} + X_{33} &\leq 77 \qquad S = 235 \\
X_{11} + X_{21} + X_{31} &= 72 \\
X_{12} + X_{22} + X_{32} &= 102 \\
X_{23} + X_{33} &= 41 \qquad D = 215 \\
X_{ij} \geq 0 \ (i,j &= 1,2,3)
\end{aligned}
$$

Notice that total supply (235) exceeds total demand (215). To solve this problem by the basic simplex algorithm, recall that it would be necessary to convert the supply inequalities to equalities by introducing a slack variable in each of the ($\leq$) constraints. It is necessary to take an equivalent step in the transportation (simplex) algorithm. Since, in this example, supply exceeds demand by 20 units, we need to create a "dummy" demand point or destination having a demand for 20 units. Had demand exceeded the supply, it would have been necessary to introduce a dummy supply point. The initial transportation table is illustrated in Table 16-38. Note that $C_{ij} = 0$ for all dummy variables introduced to "balance" the table. It should also be noted that if a particular link (X_{ij}) is not feasible (in this example, X_{13}), $C_{ij} = +M$. This will insure X_{ij} is driven out of the solution basis.

Table 16-38

	1	2	3	4(D)	SUPPLY
1	4	8	M	0	76
2	16	24	16	0	82
3	8	16	24	0	77
DEMAND	72	102	41	20	235

Once the problem is converted to standard tabular form, the next step is to identify an initial basic feasible solution. There are a number of procedures for obtaining an initial solution. The Northwest Corner Rule is a very simple, though not necessarily efficient, method. You begin in the upper-left ("northwest") corner and allocate as much as possible. In this case, 72 units can be placed in cell (1,1). This allocation uses up all of the demand in column (1), but leaves 76 - 72 = 4 units in row (1). These remaining units are allocated to (1,2). Allocating these 4 units to (1,2) leaves 98 units in column (2). However, of these, only 82 can be committed to (2,2) because the supply for the second row is limited to that amount. Allocating 82 units to (2,2) leaves 16 units in column (2). These can be assigned to (3,2) satisfying the demand of column (2) and reducing the supply in row (3) from 77 to 61. Because of the demand in column (3), only 41 units can be allocated to (3,3), satisfying the demand in that column and leaving 20 units for assignment to cell (3,4). Because the table was constructed to insure that total supply (S) and total demand (D) are equal, allocating the remaining 20 units to (3,4) closes out both row (3) and column (4).

Table 16-39

	1	2	3	4(D)	SUPPLY
1	$\boxed{4}$ 72	$\boxed{8}$ 4	$\boxed{M}$	$\boxed{0}$	~~76~~ ~~4~~
2	$\boxed{16}$	$\boxed{24}$ 82	$\boxed{16}$	$\boxed{0}$	~~82~~
3	$\boxed{8}$	$\boxed{16}$ 16	$\boxed{24}$ 41	$\boxed{0}$ 20	~~77~~ ~~61~~ ~~20~~
DEMAND	~~72~~	~~102~~ ~~98~~ ~~16~~	~~41~~	~~20~~	235

(m+n-1)
3+4-1=6

The initial basic feasible solution then is:

$$
\begin{array}{lll}
X_{11} = 72 & Z = 4(72) & = \ 288 \\
X_{12} = 4 & \quad + 8(4) & + \ \ 32 \\
X_{22} = 82 & \quad + 24(82) & + \ 1968 \\
X_{32} = 16 & \quad + 16(16) & + \ \ 256 \\
X_{33} = 41 & \quad + 24(41) & + \ \ 984 \\
X_{34} = 20 & \quad \underline{+ 0(20)} & \underline{+ \ \ \ \ 0} \\
 & & \$\ 3528
\end{array}
$$

A second procedure for determining an initial basic feasible solution is called the Least Cost Rule. In this method, we try to be "greedy" and sequentially allocate as much as possible to cells with the smallest C_{ij} value, thereby hoping to reduce the cost. Table 16-40 illustrates the initial basic feasible solution using the Least Cost method. Begin by finding the cell (i, j) with the smallest C_{ij}. In this example, cells (1, 4), (2, 4), and (3, 4) each have a $C_{ij} = 0$. You can begin by selecting one of these cells arbitrarily, or you can perhaps do a little better by looking at each option to determine which leads to the next smallest C_{ij} value. In this case, allocating 20 units to (1, 4) permits you to assign the remaining 56 units to cell (1, 1), which has a unit cost of $C_{11} = 4$. Since the next "cheapest" cell is (3, 1) with $C_{ij} = 8$, the remaining 16 units in column (1) are assigned, leaving a total of 61 in row (3). There is now a tie between (2, 3) and (3, 2) for the cell in which an allocation can be made that has the next smallest C_{ij} (=16). Since an allocation in either of these positions forces an allocation to (2, 2) where $C_{22} = 24$, you can proceed more or less arbitrarily. Assigning 61 to (3, 2) uses up the remaining supply in row (3) and reduces the demand remaining in column (2) to 41. This residual must now be assigned to (2, 2), completing the solution basis.

Table 16-40

	1	2	3	4(D)	SUPPLY
1	4 56	8	M	0 20	~~76~~ ~~56~~
2	16	24 41	16 41	0	~~82~~ ~~41~~
3	8 16	16 61	24	0	~~77~~ ~~61~~
DEMAND	~~72~~ ~~16~~	~~102~~ ~~41~~	~~41~~	~~20~~	235

The initial basic feasible solution by the Least Cost method is:

$X_{11} = 56$	$\mathcal{Z} = 4(56)$	=	224
$X_{14} = 20$	$+ \; 0(20)$	+	0
$X_{22} = 41$	$24(41)$	+	984
$X_{23} = 41$	$16(41)$	+	656
$X_{31} = 16$	$8(16)$	+	128
$X_{32} = 61$	$16(61)$	+	976
			\$ 2,968

While the Least Cost Rule is still rather conceptually unsophisticated, it does - in this case - yield an initial basic feasible solution that is \$3,528 - \$2,968 = \$560 better than that obtained using the "unthinking" Northwest Corner Rule.

The difficulty with the Least Cost decision rule is that the sequential selection of the cell with the smallest C_{ij} value may prevent, in a subsequent assignment, allocation to a cell with a lower C_{ij} value than the one selected. A procedure referred to as Vogel's Approximation Method (VAM) attempts to mitigate this problem by first computing the opportunity cost or penalty for a particular allocation. The procedure is as follows:

1. For each row and column, find the difference between the two lowest C_{ij} values. This is the "Vogel number."

2. Select the largest Vogel number and make the largest allocation possible to the cell (i, j) corresponding to the smaller of the two C_{ij} values. Ties can be broken arbitrarily.

3. After each assignment the procedure is repeated, using only the remaining rows and columns.

For the example, the Vogel computations are as follows:

ITERATION		
	ROW	
1	1	$4(C_{11}) - 0(C_{14}) = 4$
	2	$16(C_{21}/_{23}) - 0(C_{24}) = 16$ ✓
	3	$16(C_{32}) - 8(C_{31}) = 8$
	COLUMN	
	1	$8(C_{31}) - 4(C_{11}) = 4$
	2	$16(C_{32}) - 8(C_{12}) = 8$
	3	$24(C_{33}) - 16(C_{23}) = 8$
	4	$0(C_{14}/_{24}/_{34}) - 0(C_{14}/_{24}/_{34}) = 0$

Assign 20 to (2, 4)

ITERATION		
	ROW	
2	1	$8(C_{12}) - 4(C_{11}) = 4$
	2	$16(C_{21}/_{23}) - 16(C_{21}/_{23}) = 0$
	3	$16(C_{32}) - 8(C_{31}) = 8$ ✓(TIE)
	COLUMN	
	1	$8(C_{31}) - 4(C_{11}) = 4$
	2	$16(C_{32}) - 8(C_{12}) = 8$
	3	$24(C_{33}) - 16(C_{23}) = 8$

Arbitrarily select (3, 1) and assign 72

ITERATION		
	ROW	
3	1	$M(C_{13}) - 8(C_{12}) = M\checkmark$
	2	$24(C_{22}) - 16(C_{23}) = 8$
	3	$24(C_{33}) - 16(C_{32}) = 8$
	COLUMN	
	2	$16(C_{32}) - 8(C_{12}) = 8$
	3	$24(C_{33}) - 16(C_{23}) = 8$

Assign 76 to (1, 2)

ITERATION		
	ROW	
4	2	$24(C_{22}) - 16(C_{23}) = 8\checkmark$
	3	$24(C_{33}) - 16(C_{32}) = 8$
	COLUMN	
	2	$24(C_{22}) - 16(C_{32}) = 8$
	3	$24(C_{33}) - 16(C_{23}) = 8$

Arbitrarily select (2, 3) and assign 41

ITERATION		
	COLUMN	
5	2	$24(C_{22}) - 16(C_{32}) = 8$

Assign 5 to (3, 2), forcing assignment of 21 to (22)

Table 16-41

	1	2	3	4(D)	SUPPLY
1	4	8 76	M	0	~~76~~
2	16	24 21	16 41	0 20	~~82~~ ~~62~~ ~~21~~
3	8 72	16 5	24	0	~~77~~ ~~5~~
DEMAND	~~72~~	~~102~~ ~~26~~ ~~21~~	~~41~~	~~20~~	235

The initial basic feasible solution using the VAM is:

$X_{12} = 76$	$Z = 8(76)$	=	608
$X_{22} = 21$	$+ 24(21)$	=	504
$X_{23} = 41$	$+ 16(41)$	=	656
$X_{24} = 20$	$+ 0(20)$	=	0
$X_{31} = 72$	$8(72)$	=	576
$X_{32} = 5$	$16(5)$	=	80
			$ 2,424

Notice that this solution is $2,968 - 2,424 = $544 better than that obtained using the Least Cost Rule and $3,528 - 2,424 = $1,104 better than the one using the Northwest Corner Rule. When solving transportation problems manually, VAM gives a good (often optimal) initial solution at some computational expense. By contrast, the Northwest Corner and Least Cost decision rules offer a quicker initial solution, but one which is not necessarily close to the optimal solution and one which may require (many) more iterations of the transportation simplex method to reach optimality.

Once an initial basic feasible solution by either the Northwest Corner, Least Cost, or VAM approach is determined, the next step is to test that solution for optimality. Two alternative procedures are commonly used: the stepping-stone and Modified Distribution (MODI) methods.

The "stepping-stone" procedure is very similar to the basic simplex algorithm. In the simplex, the signs of the values in row (0), which represented marginal rates of contribution to the criterion (Z) for the nonbasic variables, were checked to determine whether or not the solution was optimal. In the stepping-stone procedure we do the same thing. An index representing the marginal contribution to the criterion variable value is computed for each nonbasic variable (i.e., for each empty cell). If all of these indicators are positive, no further improvement (reduction) in Z is possible. On the other hand, if one or more cell has a negative indicator, the current solution basis is not optimal. Consider again the initial basic feasible solution found by the Northwest Corner Rule and reproduced in Table 16-42. To determine if this solution is optimal, we need to examine each nonbasic variable (empty cell) and ask what the effect on Z is if one unit of this (nonbasic) variable enters the solution basis. Look, for example, at variable X_{21}. If one unit enters (2, 1), then cell (2, 2) must be reduced by one unit to keep the total at 82. However, if (2, 2) is reduced by one, then (1, 2) must be increased by one to keep the column in balance at 102. Similarly, increasing (1, 2) by one means (1, 1) must be reduced by one to keep row (1) equal to 76. Now, for this closed loop, i.e., for +(2, 1), - (2, 2), + (1, 2), - (1, 1), look at the corresponding C_{ij} values: +16 - 24 + 8 - 4 = -4. The cell indicator for nonbasic variable (2, 1) is -4, meaning that for every unit of X_{21} brought into the solution basis, Z changes by -4.

Table 16-42

	1	2	3	4(D)	SUPPLY
1	-1 4 72	+1 8 4	M	0	76
2	+1 16	-1 24 82	16	0	82
3	8	16 16	24 41	0 20	77
DEMAND	72	102	41	20	235

Let's try another one. Look at cell (2, 4), Table 16-43.

Table 16-43

	1	2	3	4(D)	SUPPLY
1	4 72	3 4	M	0	76
2	16	-1 24 82	16	+1 0	82
3	8	+1 16 16	24 41	-1 0 20	77
DEMAND	72	102	41	20	235

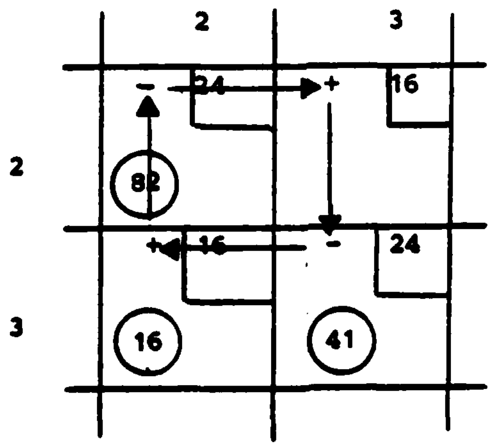
Increasing (2, 4) requires a corresponding decrease in (3, 4). Continuing, the decrease in (3, 4) requires an increase in (3, 2). Finally, the increase in (3, 2) forces a similar decrease in (2, 2). Looking at the C_{ij} values for this closed loop: + 0 - 0 + 16 - 24 = -8. The computations for each of the cell indicators are shown below:

NONBASIC VARIABLE

(1, 3) + M - 24 + 16 - 8 = + M - 16 (+)
(1, 4) + 0 - 0 + 16 - 8 = + 8
(2, 1) + 16 - 24 + 8 - 4 = - 4
(2, 3) + 16 - 24 + 16 - 24 = -16
(2, 4) + 0 - 0 + 16 - 24 = - 8
(3, 1) + 8 - 16 + 8 - 4 = -4

Since there are four negative indicators, it is apparent that this initial solution is not optimal and can be improved.

In the simplex procedure, we saw that a nonoptimal solution was improved by bringing the nonbasic variable into the basis, thereby driving one of the basic variables out. The incoming variable was that making the largest contribution to the objective. The same is true in the stepping-stone algorithm. The nonbasic variable having the most negative cell indicator is selected as the entering variable. In this example, X_{23} makes the greatest marginal reduction in Z(-16). In determining this cell indicator, we found that as X_{23} was increased, X_{33} decreased, X_{32} increased, and X_{22} decreased. The question of importance is how much X_{23} can be increased before one of the basic variables is driven out of the basis. In the general simplex algorithm, this was determined by calculating the ratio $r_i = \frac{b_i}{a_{ij*}}$. In the stepping-stone algorithm, we focus on the basic variables in the closed loop used to determine the cell indicator for the incoming variable:



As (2, 3) increases, both (3, 3) and (2, 2) decrease. When X_{23} has increased to 41, X_{33} will have decreased to 0, X_{32} will have increased to 57, and X_{22} will have decreased to 41. The leaving variable then is the

basic variable in the closed loop which has the smallest value and which decreases as the incoming variable increases in value. This new solution is illustrated in Table 16-44.

Table 16-44

	1	2	3	4(D)	SUPPLY
1	4 72	8 4	M	0	76
2	16	24 41	16 41	0	82
3	8	16 57	24	0 20	77
DEMAND	72	102	41	20	235

Checking this for optimality, we get:

NONBASIC VARIABLE

(1, 3) + M - 16 + 24 - 8 = M
(1, 4) + 0 - 0 + 16 - 8 = +8
(2, 1) + 16 - 24 + 8 - 4 = -4
(2, 4) + 0 - 0 + 16 - 24 = -8
(2, 1) + 8 - 16 + 8 - 4 = -4
(3, 3) + 24 - 16 + 24 - 16 = +16

(2, 4) enters; (3, 4) leaves; the new basis is:

$X_{11} = 72$
$X_{12} = 4$
$X_{22} = 21$
$X_{23} = 41$
$X_{24} = 20$
$X_{32} = 70$

Table 16-45 reflects this new solution. Check for optimality:

Table 16-45

	1	2	3	4(D)	SUPPLY
1	4 72	8 4	M	0	76
2	16	24 21	16 41	0 20	82
3	8	16 77	24	0	77
DEMAND	72	102	41	20	235

NONBASIC VARIABLE

(1,3) +M - 16 + 24 - 8 = M
(1,4) +0 - 0 + 24 - 8 = 16
(2,1) +16 - 24 + 8 - 4 = -4
(3,1) +8 - 16 + 8 - 4 = -4
(3,3) +24 - 16 + 24 - 16 = 16
(3,4) +0 - 0 + 24 - 16 = +8

Tie between (2,1) and (3,1) for the entering variable; select (3,1) because it permits Z to be reduced by -4(72) vs. -4(21) if (2,1) is selected. As (3,1) enters, (1,1) leaves and the new solution is:

$X_{12} = 76$

$X_{22} = 21$

$X_{23} = 41$

$X_{24} = 20$

$X_{31} = 72$

$X_{32} = 5$

Table 16-46

	1	2	3	4(D)	SUPPLY
1	4	8 76	M	0	76
2	16	24 21	16 41	0 20	82
3	8 72	16 5	24	0	77
DEMAND	72	102	41	20	235

Table 16-46 reflects this new solution.

Testing for optimality:

NONBASIC VARIABLE

(1,1) $4 - 8 + 16 - 8 = +4$
(1,3) $M - 16 + 24 - 8 = M$
(1,4) $+0 - 0 + 24 - 8 = +16$
(2,1) $+16 - 24 + 16 - 8 = 0$
(3,3) $+24 - 16 + 24 - 16 = +16$
(3,4) $+0 - 0 + 24 - 16 = +8$

The solution in Table 16-46 is optimal:

X_{ij}

$X_{12} = 76$	$Z = 8(76) = 608$
$X_{22} = 21$	$+24(21) = 504$
$X_{23} = 41$	$+16(41) = 656$
$X_{24} = 20$	$+0(20) = 0$
$X_{31} = 72$	$+8(72) = 576$
$X_{32} = 5$	$+16(5) = 80$
	\$2,424

Several points should be noted about this optimal solution. First, note that bringing (2,1) into the solution has no effect on Z. Consequently, an alternate optimum solution exists:

$X_{12} = 76$	$Z = 8(76) =$	608
$X_{21} = 21$	$+16(21) =$	+336
$X_{23} = 41$	$+16(41) =$	+656
$X_{24} = 20$	$+\ 0(20) =$	0
$X_{31} = 51$	$+\ 8(51) =$	+408
$X_{32} = 26$	$+16(26) =$	+416
		$2,424

Secondly, the optimal solution arrived at in Table 16-46 after several iterations is the same solution previously generated by the VAM procedure. This illustrates the relative efficiency of the VAM for finding an initial basic feasible (and, in this case, optimal) solution.

The MODI solution algorithm is a variation of the stepping-stone procedure that uses a somewhat more sophisticated technique for computing the indices used to test the solution for optimality and to improve nonoptimal solutions. The procedure is based on the following two relationships:

$$C_{ij} = U_i + V_j \quad \text{(For basic variables)}$$

$$I_{ij} = C_{ij} - U_i - V_j \quad \text{(For nonbasic variables)}$$

where I_{ij} represents the change in the value of the criterion variable (Z) which would result from introducing one unit of the nonbasic variable X_{ij} into the solution basis. These relationships are derived from what is known as the dual formulation of the LP model. A discussion of duality theory is not included here, but can be found in the texts referenced following this section. To illustrate the MODI procedure, the example model has been reproduced in Table 16-47. The initial basic feasible solution shown here is the one derived using the Least Cost method (see Table 16-40). Notice that the table has been modified to include a column labeled "U_i" and a row labeled "V_j".

We begin by arbitrarily setting the value of any U_i or V_j value to zero. For example, let $U_i = 0$. Now, recall that for each <u>basic</u> variable (i.e., for each "filled" cell):

$$C_{ij} = U_i + V_j$$

With this relationship, and setting one of the U_i or V_j values equal to zero, all of the remaining U_i and V_j values can be computed:

Table 16-47

	1	2	3	4(D)	SUPPLY	U_i
1	4 / 56	8	M	0 / 20	76	
2	16	24 / 41	16 / 41	0	82	
3	8 / 16	16 / 61	24	0	77	
DEMAND	72	102	41	20	235	
V_j						

$U_1 = 0$
$V_1 = C_{11} - U_1 = 4-0 = +4$
$U_3 = C_{31} - V_1 = 8-4 = +4$
$V_2 = C_{32} - U_3 = 16-4 = +12$
$U_2 = C_{22} - V_2 = 24-12 = +12$
$V_3 = C_{23} - U_2 = 16-12 = +4$
$V_4 = C_{14} - U_1 = 0-0 = 0$

Having computed the U_i and V_j values using the relationships $C_{ij} = U_i + V_j$ for the basic cells, we can compute improvement indices for each nonbasic cell as $I_{ij} = C_{ij} - U_i - V_j$:

NONBASIC VARIABLE

(1,2) $I_{12} = 8 - 0 - 12 = -4$
(1,3) $I_{13} = M - 0 - 4 = +M$
(2,1) $I_{21} = 16 - 12 - 4 = 0$
(2,4) $I_{24} = 0 - 12 - 0 = -12$
(3,3) $I_{33} = 24 - 4 - 4 = +16$
(3,4) $I_{34} = 0 - 4 - 0 = -4$

Since three of the indices are negative, the solution can be improved (reduced). Since I_{24} is the most negative, X_{24} is selected as the entering variable. The leaving variable is found as it was in the stepping-stone procedure by finding the basic variable that is driven to zero first as the entering variable (X_{24}) increases. The loop for X_{24} is: +(2,4) - (2,2) + (3,2) - (3,1) + (1,1) - (1,4) as shown in Table 16-48.

Table 16-48

	1	2	3	4(D)	SUPPLY	U_i
1	$\boxed{4}$ 56 +	-4 $\boxed{8}$	+M $\boxed{M}$	0 20 -	76	0
2	0 $\boxed{16}$	$\boxed{24}$ 41 -	$\boxed{16}$ 41	-12 $\boxed{0}$ +	82	12
3	$\boxed{8}$ 16 -	$\boxed{16}$ 61 +	+16 $\boxed{24}$	-4 $\boxed{0}$	77	4
DEMAND	72	102	41	20	235	
V_j	4	12	4	0		

As X_{24} enters, X_{31} will be driven out of the basis when $X_{24} = 16$. The new basis will be:

$X_{11} = 72$, $X_{14} = 4$, $X_{22} = 25$, $X_{23} = 41$, $X_{24} = 16$, $X_{32} = 77$ as shown in Table 16-49.

Table 16-49

	1	2	3	4(D)	SUPPLY	Ui
1	4; (72)	+ 8	M	− 0; (4)	76	0
2	16	− 24; (25)	16; (41)	+ 0; (16)	82	0
3	8	16; (77)	24	0	77	-8
DEMAND	72	102	41	20	235	
	4	24	16	0		

To determine if this solution is optimal, we again compute the U_i and V_j values:

$U_1 = 0$
$V_4 = C_{14} - U_1 = 0-0 = 0$
$U_2 = C_{24} - V_4 = 0-0 = 0$
$V_3 = C_{23} - U_2 = 16-0 = +16$
$V_2 = C_{22} - U_2 = 24-0 = +24$
$U_3 = C_{32} - V_2 = 16-24 = -8$
$V_1 = C_{11} - U_1 = 4-0 = +4$

Computing the indices for nonbasic cells:

$I_{12} = C_{12} - U_1 - V_2 = 8 - 0 - 24 = -16$
$I_{13} = C_{13} - U_1 - V_3 = M - 0 - 16 = +M$
$I_{21} = C_{21} - U_2 - V_1 = 16 - 0 - 4 = +12$
$I_{31} = C_{31} - U_3 - V_1 = 8 - (-8) - 4 = +12$
$I_{33} = C_{33} - U_3 - V_3 = 24 - (-8) - 16 = +16$
$I_{34} = C_{34} - U_3 - V_4 = 0 - (-8) - 0 = +18$

Solution is not optimal; X_{12} enters and X_{14} leaves. The new solution is: $X_{11} = 72$, $X_{12} = 4$, $X_{22} = 21$, $X_{23} = 41$, $X_{24} = 20$, $X_{32} = 77$.

Table 16-50

	1	2	3	4(D)	SUPPLY	Ui
1	− [4] 72	+ [8] 4	[M]	[0]	76	0
2	[16]	[24] 21	[16] 41	[0] 20	82	16
3	+ [8]	− [16] 77	[24]	[0]	77	8
DEMAND	72	102	41	20	235	
Vj	4	8	0	−16		

Checking for optimality, we again compute new U_i and V_j values:

$U_1 = 0$
$V_1 = C_{11} - U_1 = 4 - 0 = 4$
$V_2 = C_{12} - U_1 = 8 - 0 = 8$
$U_2 = C_{22} - V_2 = 24 - 8 = 16$
$V_3 = C_{23} - U_2 = 16 - 16 = 0$
$V_4 = C_{24} - U_2 = 0 - 16 = -16$
$U_3 = C_{32} - V_2 = 16 - 8 = 8$

Computing the indices for the nonbasic cells:

$I_{13} = C_{13} - U_1 - V_3 = M - 0 - 0 = M$
$I_{14} = C_{14} - U_1 - V_4 = 0 - 0 -(-16) = +16$
$I_{21} = C_{21} - U_2 - V_1 = 16 - 16 - 4 = -4$
$I_{31} = C_{31} - U_3 - V_1 = 8 - 8 - 4 = -4$
$I_{33} = C_{33} - U_3 - V_3 = 24 - 8 - 0 = +16$
$I_{34} = C_{34} - U_3 - V_4 = 0 - 8 -(-16) = +8$

Solution is not optimal; there is a tie between X_{21} and X_{31} as the entering variable. Select X_{31} to enter. As X_{31} enters, X_{11} leaves and the new solution basis is: $X_{12} = 76$, $X_{22} = 21$, $X_{23} = 41$, $X_{24} = 20$, $X_{31} = 72$, $X_{31} = 5$. This solution is shown in Table 16-51

Table 16-51

	1	2	3	4(D)	Supply	U_i
1	4	8 76	M	0	76	0
2	16	24 21	16 41	0 20	82	16
3	8 72	16 5	24	0	77	8
DEMAND	72	102	41	20	235	
V_j	0	8	0	-16		

Again, checking for optimality, compute U_i and V_j values:

$U_1 = 0$
$V_2 = C_{12} - U_1 = 8 - 0 = 8$
$U_2 = C_{22} - V_2 = 24 - 8 = 16$
$V_3 = C_{23} - U_2 = 16 - 16 = 0$
$V_4 = C_{24} - U_2 = 0 - 16 = -16$
$U_3 = C_{32} - V_2 = 16 - 8 = 8$
$V_4 = C_{31} - U_3 = 8 - 8 = 0$

Computing the indices for the nonbasic variables:

$I_{11} = C_{11} - U_1 - V_1 = 4 - 0 - 0 = +4$
$I_{13} = C_{13} - U_1 - V_3 = M - 0 - 0 = +M$
$I_{14} = C_{14} - U_1 - V_4 = 0 - 0 -(-16) = +16$
$I_{21} = C_{21} - U_2 - V_1 = 16 - 16 - 0 = 0$
$I_{33} = C_{33} - U_3 - V_3 = 24 - 8 - 0 = +16$
$I_{34} = C_{34} - U_3 - V_4 = 0 - 8 -(-16) = +8$

The solution of Table 16-51 is optimal.

$$Z = 8(76) + 24(21) + 16(41) + 0(20) + 8(72) + 16(5)$$
$$Z = 608 + 504 + 656 + 0 + 576 + 80$$
$$Z = \underline{\$2,424}$$

In the process of developing the initial or a subsequent basic feasible solution, it is not uncommon to encounter the abberation of degeneracy. In the transportation model, a solution is degenerate when there are fewer than (m+n-1) basic variables. When this condition occurs, it is no longer possible to compute the improvement indices. Consider, for example, the transportation model illustrated in Table 16-52.

Table 16-52

	1	2	3	SUPPLY
1	4 35	8 20	8	~~55~~ ~~20~~
2	16	24 25	16	~~25~~
3	8	16	24 35	35
DEMAND	~~35~~	~~45~~ ~~25~~	35	115

(m+n-1)
3+3-1=5

Only 4 basic cells: initial solution is degenerate.

Using the Northwest Corner Rule to derive an initial basic feasible solution, we assign 35 to (1,1), closing out column (1) and leaving 20 in row (1). These 20 are assigned to (1,2), closing out row (1) and leaving 25 in column (2). However, when we next allocate those remaining 25 units to (2,2), it simultaneously closes out both row (2) and column (2). Consequently, there are only 4 basic variables rather than the m(3)+n(3)-1=5 required. The initial basic feasible solution is degenerate. Suppose you had not recognized this condition and were attempting to compute improvement indices for the nonbasic cells (using either the stepping-stone or MODI procedures). For cells (1,3), (2,3), (3,1), and (3,2), it is not possible to form a closed loop.

Fortunately, this is not a difficult condition to resolve. Simply introduce a "0" allocation (sometimes represented by "ε") to one of the cells. While there is usually a great deal of flexibility in selecting the cell for the 0 allocation, the general procedure, when using the Northwest Corner Rule, is to assign it to the cell that maintains the "stair-step" pattern as shown in Table 16-53.

Table 16-53

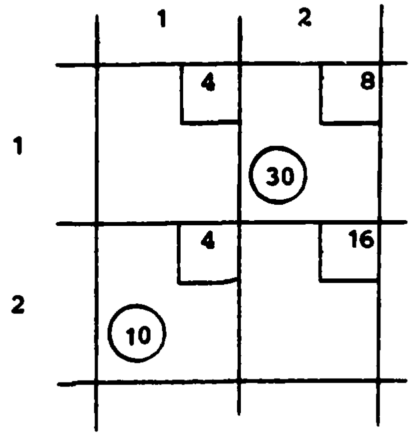
	1	2	3	SUPPLY
1	4 35	8 20	8	~~55~~ ~~20~~
2	16	24 25	16 0	~~25~~
3	8	16	24 35	35
DEMAND	~~35~~	~~45~~ ~~25~~	35	

The problem can now be solved, treating the variable having the ε or 0 allocation like any other basic variable. The 0 allocation is simply a computational device. It has no significance in the problem.

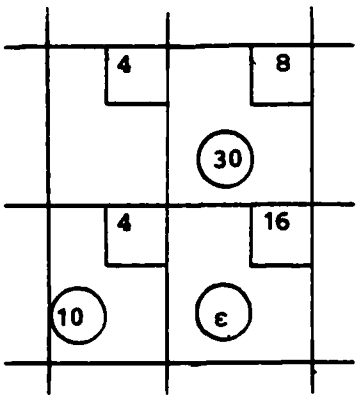
Degeneracy can also surface after the initial solution has been determined. Consider the partial transportation table shown below:

	1	2
1	4 10	8 20
2	-8 4	16 10

Nonbasic variable X_{21} has a negative cell indicator. The solution is not optimal. As X_{21} enters, however, both X_{11} and X_{22} go to 0 simultaneously (a tie for the leaving variable).



As in the case of the degenerate initial basic feasible solution, we correct this situation by introducing an ε or 0 allocation into one of those cells simultaneously driven out of the basis and proceed with the algorithm, treating the cell with the ε entry like any other basic variable:



The Assignment Model

Just as the transportation model is a special case of the general LP formulation, the assignment model is a special case of the transportation model (and, therefore, a special case of the LP model). The unique structure of the model permits solution by techniques which are computationally more efficient than either the transportation or simplex algorithms.

In the assignment problem, there are m=n supply and demand points. Unlike the transportation problem, however, all of the commodity available at a supply point (i) must be allocated to one of the demand points (j). In the prototype problem, the objective is to assign n individuals to n jobs in such a way as to minimize cost, maximize efficiency, etc. Like the transportation model, the assignment formulation is important and useful because it is applicable to a wide variety of managerial problems.

General System Analysis

System Components:

1. Criterion Variable (Z): Total cost to distribute all of the commodity available at each of n supply points to n demand points; the objective is to minimize Z.

2. Decision Variable (X_{ij}): X_{ij} =1 if the supply at i is assigned to demand point j and 0 if it is not (i,j=1,2,...,n).

3. Environmental Variables:

S_i = amount of commodity assigned from supply point i (S_i = 1, i=1,2,...,n)

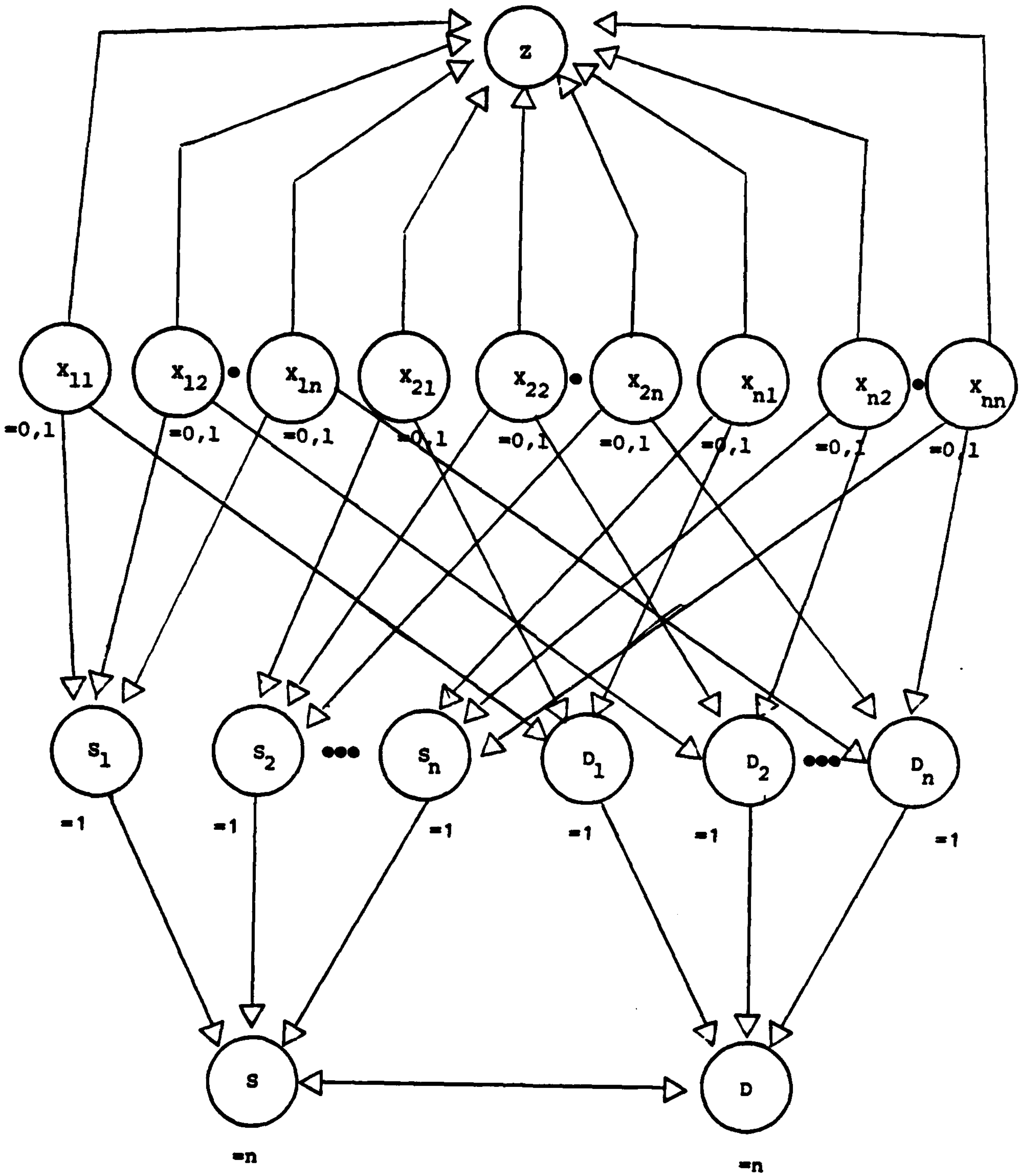
D_j = amount of commodity assigned to demand point j (D_j = 1, j=1,2,...,n)

S = total supply at all sources

D = total demand at all destinations

System Relationships

Figure 16-20



Mathematical Formulation

$$\text{minimize:} \quad Z = \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} X_{ij}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} X_{ij} = 1 \quad (i = 1,2,\ldots,n)$$

$$\sum_{i=1}^{n} X_{ij} = 1 \quad (j = 1,2,\ldots,n)$$

$$X_{ij} = 0,1$$

Note that since, in the assignment model, m=n and $S_i = d_j = 1$, total supply equals total demand.

Formulation Examples

Example 16-2, is a "classical" assignment problem involving the assignment of people to tasks.

EXAMPLE 16-2

Three new machines of different types have been purchased for installation and use in one of the shops at an Air Logistics Center. There are four available locations in the shop at which one of the machines could be installed. However, some of these locations are more desirable than others for particular machines because of their proximity to various work centers. The objective is to assign the machines to the potential locations in such a way as to minimize the total cost of materials handling. The estimated cost per unit time of materials handling involving each of the machines is given in Table 16-54. Note that machine 2 cannot be assigned to location 2:

Table 16-54

		LOCATION			
		1	2	3	4
MACHINE	1	10	15	20	12
	2	14	X	12	8
	3	6	18	10	15

How should the machines be located?

System Analysis

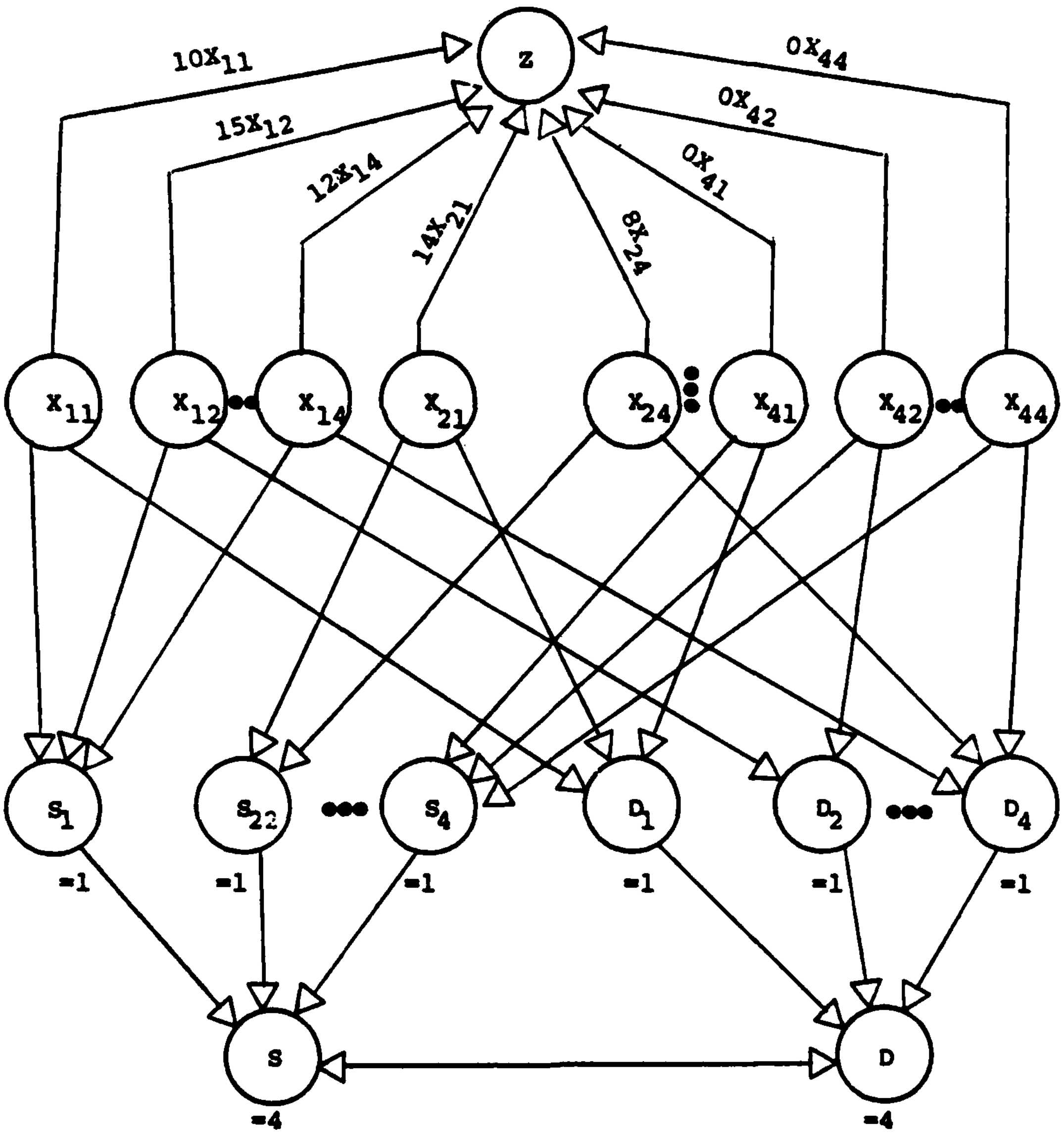
System Components:

1. Criterion Variable (Z): Total cost per unit time; objective is to minimize Z.

2. Decision Variable $(X_{ij})=1$ if machine i is assigned to location j; 0 otherwise. (Note: to formulate this as an assignment model with m=n, it is necessary to introduce a "dummy" machine to fill the extra location).

3. Environmental Variables: S_i = number of locations assigned to machine i (S_i = 1; i=1,2,...,n). D_j = number of machines assigned to location j (D_j = 1; j=1,2,...,n).

Figure 16-21



Mathematical (LP) Model

$$\begin{aligned}\text{minimize:}\quad Z = {} & 10X_{11} + 15X_{12} + 20X_{13} + 12X_{14} + 14X_{21} + 12X_{23} \\ & + 8X_{24} + 6X_{31} + 18X_{32} + 10X_{33} + 15X_{34} + 0X_{41} \\ & + 0X_{42} + 0X_{43} + 0X_{44}\end{aligned}$$

$$\begin{aligned}\text{s.t.}\quad & X_{11} + X_{12} + X_{13} + X_{14} = 1 \\ & X_{21} + X_{22} + X_{23} + X_{24} = 1 \\ & X_{31} + X_{32} + X_{33} + X_{34} = 1 \\ & X_{41} + X_{42} + X_{43} + X_{44} = 1 \\ \\ & X_{11} + X_{21} + X_{31} + X_{41} = 1 \\ & X_{12} + X_{22} + X_{32} + X_{42} = 1 \\ & X_{13} + X_{23} + X_{33} + X_{43} = 1 \\ & X_{14} + X_{24} + X_{34} + X_{44} = 1\end{aligned}$$

$$X_{ij} = 0,1 \quad (i,j = 1,2,\ldots,4)$$

Solution Techniques

As a special case of the transportation model, the assignment problem can be solved with either the transportation algorithm or the more general simplex algorithm. In addition, a more computationally efficient technique, known as the Hungarian Method, has been devised for solution of assignment problems. Basically, this procedure employs the concept of opportunity costs and can be illustrated with the example presented in the previous section:

The procedure begins by subtracting from each row and column of the cost table the smallest value in that row or column.

1.

	1	2	3	4	Smallest Row Value		
1	10	15	20	12	10		
2	14	M	12	8	8		
3	6	18	10	15	6		
4	0	0	0	0	0	=	24

Subtracting these minimum row values:

	1	2	3	4
1	0	8	10	2
2	6	M	4	0
3	0	12	4	9
4	0	0	0	0

Now, subtracting the minimum column values gives the same table because there is at least one 0 in each column.

In this algorithm, assignments are made, if possible, to cells with "0" values. After step (1) above, there will be at least one "0" in each row and column. A convenient method exists for determining if the solution is optimal. Attempt to "cover" all of the zero values by drawing vertical and/or horizontal straight lines. If the minimum number of lines required to cover all zeros equals n, the number of rows and column, the solution is optimal. If the number of lines required to cover all zeros is less than n, the optimal solution cannot be made. In this example, an optimal solution is not yet possible as all of the zeros can be covered by a minimum of three lines:

	1	2	3	4	
1	0	5	10	2	
2	6	M	4	0	
3	0	12	4	9	4
4	0	0	0	0	

To improve the solution, find the smallest "uncovered" value and:

1. subtract it from the other uncovered values, and

2. add it to those values at the intersections of two lines.

In this example, 4 is the smallest uncovered value. Subtracting it from the other uncovered values and adding it to the intersection values, i.e., (4,1) and (4,4), gives the following revised table:

	1	2	3	4
1	0	1	6	2
2	6	M	0	0
3	0	8	0	9
4	4	0	0	4

Now, returning to step (2), we attempt to cover all the zeros and see that a minimum of 4 lines is required to cover all the zeros. Since n=4, an optimal assignment is possible:

	1	2	3	4
1	0	1	6	2
2	6	M	0	0
3	0	8	0	9
4	4	0	0	4

To make the optimal assignment, first make the assignments to the rows or columns for which there is only one zero:

	1	2	3	4
1	$\boxed{0}$	1	6	2
2	6	M	0	$\boxed{0}$
3	0	8	0	9
4	4	$\boxed{0}$	0	4

Row 1, Column 2, and Column 4 have only one zero each. Therefore, these assignments must be made, as indicated by the square drawn around the 0. Since each row and column can have only one $X_{ij}=1$, the optimal assignment is:

$$
\begin{array}{ll}
X_{11} = 1 & C_{11} = 10 \\
X_{24} = 1 & C_{24} = 8 \\
X_{33} = 1 & C_{33} = 10 \\
X_{42} = 1 & C_{42} = 0 \\
\hline
 & Z = 28
\end{array}
$$

Note that the sum of all values subtracted from the table (24 + 4) is equal to Z.

REFERENCES

1. Anderson, D. R., D. J. Sweeney, and T. A. Williams. An Introduction to Management Science: Quantitative Approaches to Decision Making. St Paul, MN: West Publishing Co., 1976.

2. Budnick, F. S., R. Mojena, and T. E. Vollmann. Principles of Operations Research for Management. Homewood, IL: Richard D. Irwin, Inc., 1979.

3. Daellenbach, H. G., and J. A. George. Introduction to Operations Research Techniques. Boston, MA: Allyn and Bacon, Inc., 1978.

4. Heinze, D. Management Science: Introductory Concepts and Applications. Cincinnati, OH: South-Western Publishing Co., 1978.

5. Hillier, F. S., and G. J. Lieberman. Operations Research. 2d ed. San Francisco, CA: Holden-Day, Inc., 1974.

6. Levin, R. I., and C. A. Kirkpatrick. Quantitative Approaches to Management. 4th ed. New York: McGraw-Hill Book Company, 1978.

7. Shore, B. Quantitative Methods for Business Decisions: Text and Cases. New York: McGraw-Hill Book Company, 1978.

8. Thompson, G. E. Management Science: An Introduction to Modern Quantitative Analysis and Decision Making. New York: McGraw-Hill Book Company, 1976.

9. Turban, E., and J. R. Meredith. Fundamentals of Management Science. Dallas, TX: Business Publications, Inc., 1977.

10. Wagner, H. Principles of Management Science: With Applications to Executive Decisions. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1976.

CHAPTER 17

NETWORK ANALYSIS

INTRODUCTION

In the discussion of linear programming, we (hopefully) demonstrated the importance and utility of visualizing and analyzing a decision situation as a system. In describing a system, we suggested that any system is comprised of a set of relevant dependent and independent variables linked by relationships. An important system characteristic, implied but not emphasized, was that such systems can be interpreted as a network through which some element flows. The element flowing through a network system may be virtually anything, either physical or nonphysical, e.g., people, money, time, information, equipment, vehicles, social influence, etc. The concept of network systems and the study of such systems (network analysis) has played an important role in such disciplines as electrical engineering, communications, computer science, transportation engineering, and cybernetics. More recently, network systems and their analysis have surfaced as an important tool for managers in a variety of organizational settings. Perhaps most notably, network analysis has emerged as a fundamental technique for managing the timely and effective accomplishment of projects, particularly in the research and development and the construction areas. In addition, network analysis is now being used as a basic tool for studying social, economic, and political systems.

In this section, basic network concepts are introduced. These concepts are used in a number of algorithms which have been developed to solve network-type problems which frequently occur in a variety of managerial decision situations. Specifically, we will discuss the following network models:

a. Shortest Route Problem

b. Minimal Spanning Tree Problem

c. Maximal Flow Problem

d. Closed Circuit Problem

This section concludes with a discussion of project (system) management and the use of such techniques as the Critical Path Method (CPM) and the Program Evaluation and Review Technique (PERT).

Basic Network Concepts and Definitions

To facilitate the discussion of network systems and the analysis of such systems, it is necessary to introduce a number of basic concepts and definitions.

1. **BRANCH**: A line segment connecting two distinct end points or stations.

(a) **NODE**: The end point of a branch. Also called a vertex or junction point

(b) **ARC**: The line segment of a branch connecting the two bounding nodes. Also known as edges or links.

Figure 17-1. BRANCH

(c) **NODE INDEX**: Each node has a unique label, i.e., a letter or number. A specific branch is identified by the labels on its nodes (i,j).

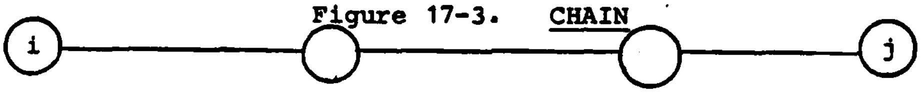
(d) **DIRECTED BRANCH**: A branch in which the direction of flow is specified. The direction of flow is indicated with an arrow:

Figure 17-2. DIRECTED BRANCH

(e) **CAPACITATED BRANCH**: A branch in which the magnitude of flow is restricted to some limiting value.

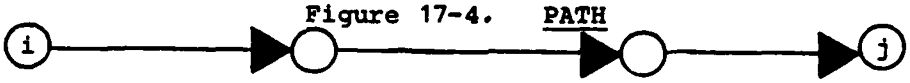
2. **CHAIN**: A sequence of branches connecting node i with node j, such that each node in the sequence is encountered only once:

Figure 17-3. CHAIN



3. **PATH**: A chain in which the direction of flow is specified:

Figure 17-4. PATH

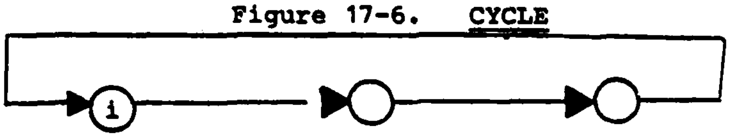


4. **LOOP**: A branch in which the two nodes are coincident:

Figure 17-5. LOOP

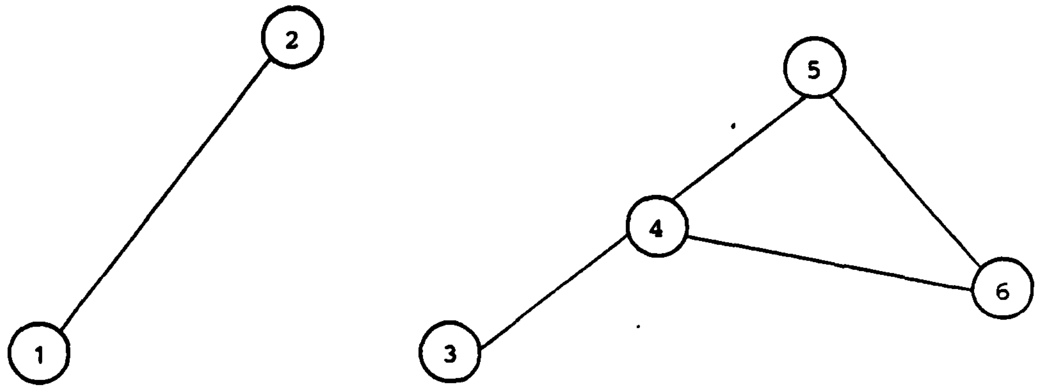
5. **CYCLE**: A chain in which the two terminal nodes are coincident:

Figure 17-6. CYCLE



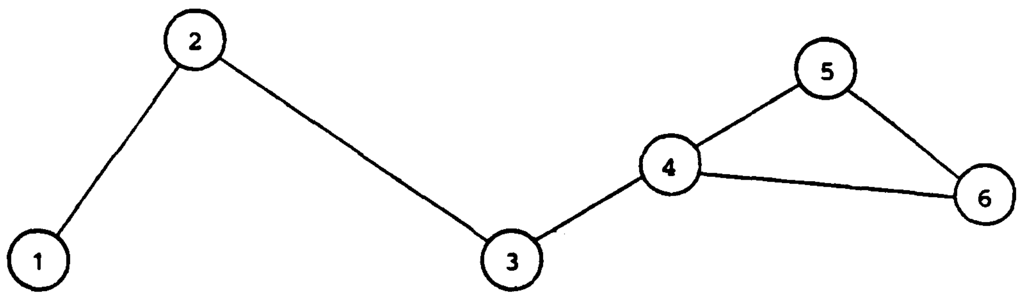
6. GRAPH: A set of points or stations, certain pairs of which are connected by arcs to form branches:

Figure 17-7. GRAPH



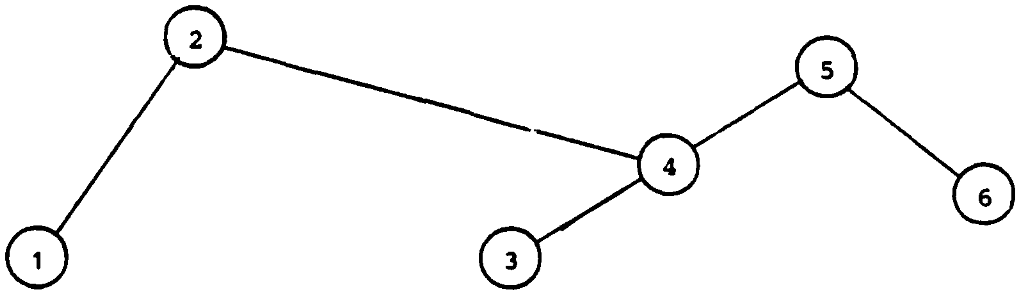
(a) CONNECTED GRAPH: A graph in which it is possible to reach any node j from (directly or indirectly) any other node i:

Figure 17-8. CONNECTED GRAPH



7. TREE: A connected graph containing no cycles. A tree of n nodes has exactly (n-1) arcs:

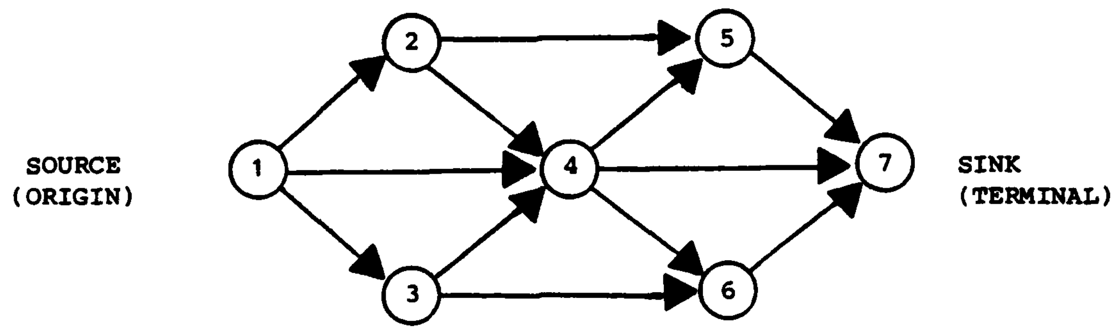
Figure 17-9. TREE



8. NETWORK: A connected graph with a flow (or potential for flow) of some type in its branches.

9. <u>DIRECTED NETWORK</u>: A network in which the direction of flow is specified for all branches:

Figure 17-10. <u>DIRECTED NETWORK</u>



10. <u>SOURCE NODE</u>: A node in a network from which all flow moves away in each incident branch. Also called an <u>ORIGIN NODE</u>.

11. <u>SINK NODE</u>: A node in a network toward which the flow moves in all incident links or branches. Also called a <u>TEMINAL NODE</u> or <u>DESTINATION NODE</u>.

12. <u>INCIDENCE MATRIX</u>: A matrix that provides a complete statement or summary of the graph or network structure.

Figure 17-11. <u>INCIDENCE MATRIX</u>

<u>TO (j)</u>

<u>FROM (i)</u>		1	2	3	4	5	6	7
	1		X	X	X			
	2				X	X		
	3				X		X	
	4					X	X	X
	5							X
	6							X
	7							

If the network is capacitated or the branches have some specific values associated with them, these values can be entered in the matrix. The incidence matrix is important because this is the vehicle used to enter the network structure in a computer program.

13. OTHER NETWORK CONCEPTS:

1. Network diagrams are generally schematic only and not drawn to scale.

2. If the network is a model of a dynamic system, you should specify the system state life, i.e., the period over which the network structure holds or is expected to hold.

3. Values associated with network branches (and/or nodes) may be either deterministic or stochastic, depending on the nature of the system being modeled.

4. There are usually no "line losses" in the system, i.e., there is conservation of flow. The amount of flow entering a node must equal the amount leaving the node. However, in multi-period models, flow may be "stored" for one or more periods.

17.1 NETWORK SYSTEM MODELS

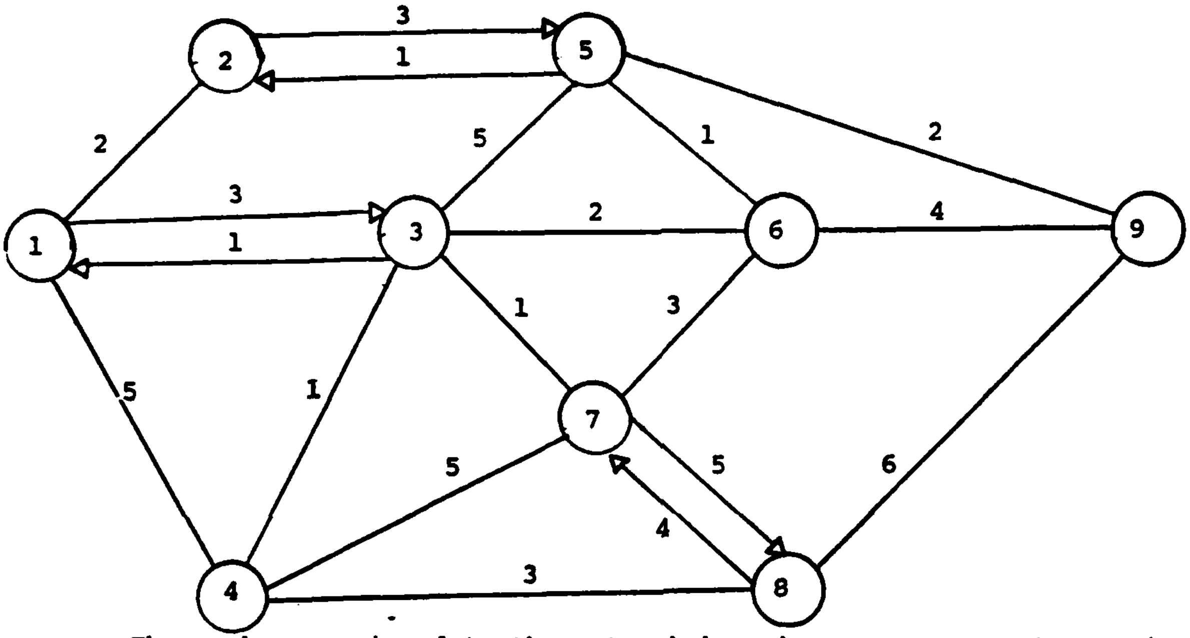
We will now look at the formulation and analytical solution of four rather "classical" network analysis problems. In the minimal path or shortest route problem the objective is to find the shortest (or, alternatively, least expensive) route through a network that connects some source and destination. The minimal spanning tree problem is a variation of the shortest route model. Here the problem is to link up a set of nodes with branches in such a way as to minimize the total length, cost, time, etc. In the maximal flow problem we are concerned with a capacitated network and want to know what maximum flow rate can be carried through the network. The final problem considered in this section, the closed circuit or "traveling salesman" problem, deals with a network in which we want to know the shortest route through the network that encounters each node once and only once and then returns to the starting point. For each model, the discussion will include a brief description of the problem and one of the available solution algorithms for that particular formulation. An example is also presented to illustrate application of the solution algorithm. It should be pointed out that each of these problems can be formulated as a mathematical model. However, these models are, in some cases, rather complex and beyond the scope of a handbook discussion. A number of excellent references on network analysis are available for those interested in a more complete and detailed treatment of the subject.

MINIMAL PATH/SHORTEST ROUTE PROBLEM

Problem Description

Given a (directed or undirected) network in which each branch has some assigned magnitude, find the path from some specified node (i) to some other specified node (j) having the smallest total magnitude, i.e., the minimal path or shortest route between (i) and (j). For example, find the minimal path between node (1) and node (9) in Figure 17-12.

Figure 17-12



The numbers assigned to the network branches may represent a variety of system characteristics, e.g., time, distance, cost, risk or other measures of utility. If, for example, the numbers in the above network represent cost, we would be interested in finding the least expensive route between two specified nodes, e.g., node (1) and node (9). Note that there is no requirement for the "flow" between two nodes to be the same, i.e., $C_{ij} \neq C_{ji}$.

Solution Techniques

A number of general and specialized techniques have been developed to solve minimal path problems, e.g.:

1. mathematical programming techniques:

    (a) linear programming

    (b) integer programming

    (c) dynamic programming

2. specialized network algorithms:

   (a) graphical method

   (b) labeling method

   (c) matrix method

Mathematical programming approaches will handle reasonably large problems but don't, in general, take advantage of the special mathematical structure of the problem. The graphical algorithms are easy to apply but are practically limited to small networks. The labeling and matrix methods can handle relatively large networks in a more efficient manner than can the mathematical programming techniques. In addition, these two specialized algorithms can be readily programmed on the computer.

EXAMPLE 17-1

Labeling Method For Solution Of The Minimal Path Problem

1. Construct the network diagram, labeling all nodes and arcs as shown in Figure 17-12.

2. Construct a table having one column for each node:

Table 17-1

1	2	3	4	5	6	7	8	9

3. In the column under each node index, list all branches leading out of that node, together with the corresponding branch value, i.e., distance, time, cost, etc. Branches leading into the origin or away from the destination need not be listed. The branches and their associated values should

be listed in order of increasing branch value, i.e., the smallest value first, etc.:

Table 17-2

1	2	3	4	5	6	7	8	9
1,2-2	2,5-3	3,4-1	4,3-1	5,2-1	6,5-1	7,3-1	8,4-3	
1,3-3		3,7-1	4,8-3	5,6-1	6,3-2	7,6-3	8,7-4	
1,4-5		3,6-2	4,7-5	5,9-2	6,7-3	7,4-5	8,9-6	
		3,5-5		5,3-5	6,9-4	7,8-5		

4. "Label" the origin node (1) by entering a "0" above the node index (column heading). This indicates the minimal distance from the origin to this node. Since this node is the origin, the minimal distance must be 0. At this point then, only node 1 is labeled; all other nodes are unlabeled:

Table 17-3

0								
1	2	3	4	5	6	7	8	9

5. For each unlabeled node that is connected directly to the labeled node (at this point, nodes 2, 3, and 4), compute the minimal path from the origin by adding the branch value to the minimal path value previously found for the labeled node (0 for the origin):

(1,2):	$0 + 2 = 2$	(minimal)
(1,3)	$0 + 3 = 3$	
(1,4)	$0 + 5 = 5$	

6. Select the unlabeled node that has the smallest total path value, i.e., the unlabeled node that is closest to the origin. Note that if the above procedures have been followed correctly, this will be the first entry in the origin column. As determined in the previous step, for this case, node 2 is closest to the origin.

7. Label this closest node (2) with the minimal path value found in step 5 (i.e., 2). Now nodes 1 and 2 are labeled.

Table 17-4

0	2							
1	2	3	4	5	6	7	8	9

8. In the column corresponding to the origin, circle the branch to the selected node (i.e., "1,2-2"). This indicates that the minimal path to node 2 from the origin node is 2 units and that the route is branch (1,2).

9. Since the minimal path to node 2 has now been found (recognized by the fact that node 2 has been labeled), all other branches leading into this node may be eliminated from further consideration. Consequently, cross out all such branches in the table (i.e., all branches (i,j) in which j=2). In this case, branch (5,2) is crossed out:

Table 17-5

0	2							
1	2	3	4	5	6	7	8	9
1,2-2	2,5-3	3,4-1	4,3-1	~~5,2-1~~	6,5-1	7,3-1	8,4-3	
1,3-3		3,7-1	4,8-3	5,6-1	6,3-2	7,6-3	8,7-4	
1,4-5		3,6-2	4,7-5	5,9-2	6,7-3	7,4-5	8,9-6	
		3,5-5		5,3-5	6,9-4	7,8-5		

10. For <u>each</u> labeled node (now 1 and 2), add the minimal path value for that node to the smallest branch value in that column. The node corresponding to the smallest of these sums is the next node to be labeled as nearest the origin. If two or more nodes are tied for the smallest sum, then all such nodes are labeled:

(1,3):	0 + 3 = 3	(minimum)
(2,5)	2 + 3 = 5	

11. Label the selected node (node 3) with the minimal path value computed in the previous step (i.e., 3). Circle the selected branch (1,3) and cross out all other branches leading into the selected node:

Table 17-6

0	2	3						
1	2	3	4	5	6	7	8	9
1,2-2	2,5-3	3,4-1	~~4,3-1~~	~~5,3-1~~	6,5-1	~~7,3-1~~	8,4-3	
1,3-3		3,7-1	4,8-3	5,6-1	~~6,3-2~~	7,6-3	8,7-4	
1,4-5		3,6-2	4,7-5	5,9-2	6,7-3	7,4-5	8,9-6	
		3,5-5		~~5,3-5~~	6,9-4	7,8-5		

12. Repeat steps 10 and 11 until the terminal node (in this case node 9) has been labeled:

(1,4): 0 + 5 = 5

(2,5): 2 + 3 = 5

(3,4): 3 + 1 = 4

(Tie: Label both 4 and 7)

(3,7): 3 + 1 = 4

<u>Circle (3,4) and (3,7); cross out all (i,4) and (i,7) branches.</u>

(2,5): 2 + 3 = 5

(Tie: Label both 5 and 6)

(3,6): 3 + 2 = 5

(4,8): 4 + 3 = 7

(7,6): 4 + 3 = 7

<u>Circle (2,5) and (3,6); cross out all (i,5) and (i,6) branches.</u>

(4,8):	4	+ 3	= 7	(Tie: Label both 8 and 9)
(5,9):	5	+ 2	= 7	
(6,9):	5	+ 4	= 9	
(7,8):	4	+ 5	= 9	

Circle (4,8) and (5,9); cross out all (i,8) and (i,9) branches.

When node 9 has been labeled, we have found the minimal path from node 1 to node 9. To trace the path, work "backwards" through the circled nodes in the table, i.e., 9-5-2-1. Therefore, the minimal path value of 7 units is found along path 1-2-5-9. The completed table is as follows:

Table 17-7

0	2	3	4	5	6	7	8	9
1	2	3	4	5	6	7	8	9
1,2-2	2,5-3	3,4-1	~~4,[illegible]-1~~	~~5,[illegible]-1~~	~~6,[illegible]-1~~	~~7,[illegible]-1~~	~~8,[illegible]-3~~	
1,3-3		3,7-1	4,8-3	~~5,[illegible]-1~~	~~6,[illegible]-2~~	~~7,[illegible]-3~~	~~8,7-4~~	
~~1,[illegible]-5~~		3,6-2	~~4,[illegible]-5~~	5,9-2	~~6,[illegible]-3~~	~~7,[illegible]-5~~	~~8,[illegible]-6~~	
		~~3,[illegible]-5~~		~~5,[illegible]-5~~	~~6,[illegible]-4~~	~~7,[illegible]-5~~		

MINIMAL SPANNING TREE PROBLEM

Problem Description

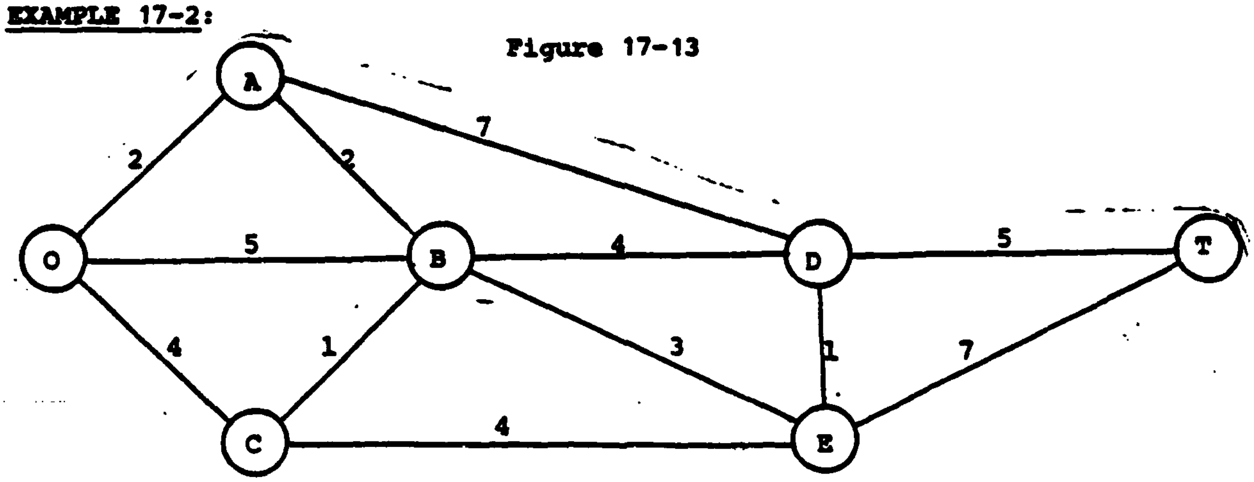
The minimal spanning tree problem is a variation of the minimal path/shortest route problem. Given a (directed or undirected) network in which each branch has some assigned magnitude or value, find the tree having the minimal total value, i.e., select the branches having the smallest total value while providing a direct or indirect route between each pair of nodes.

Solution Technique

1. Draw the network diagram, labeling all nodes and arcs.
2. Arbitrarily select any node and connect it to the nearest distinct node.
3. Identify the unconnected node that is nearest to a connected node and then connect these two nodes. Repeat this step until all nodes have been connected.

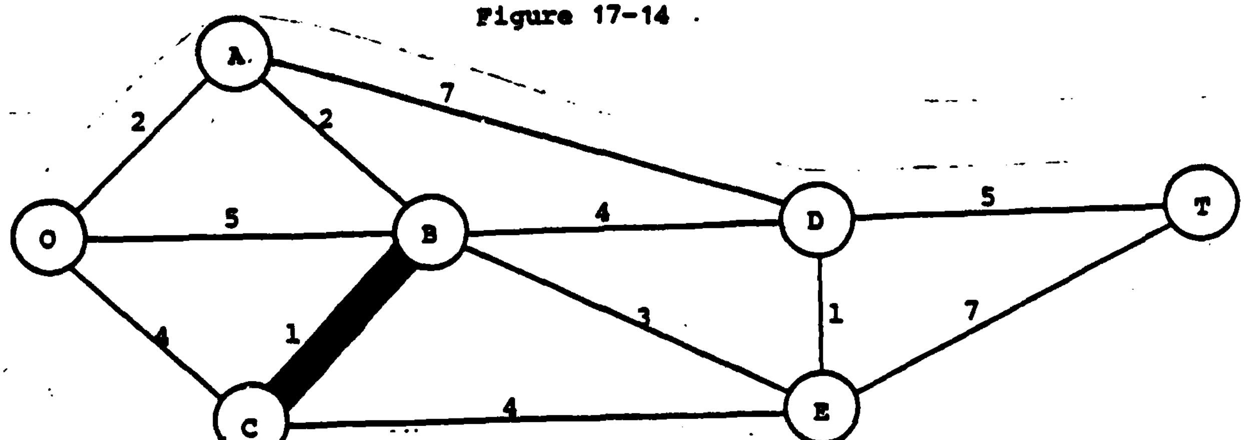
EXAMPLE 17-2:

Figure 17-13



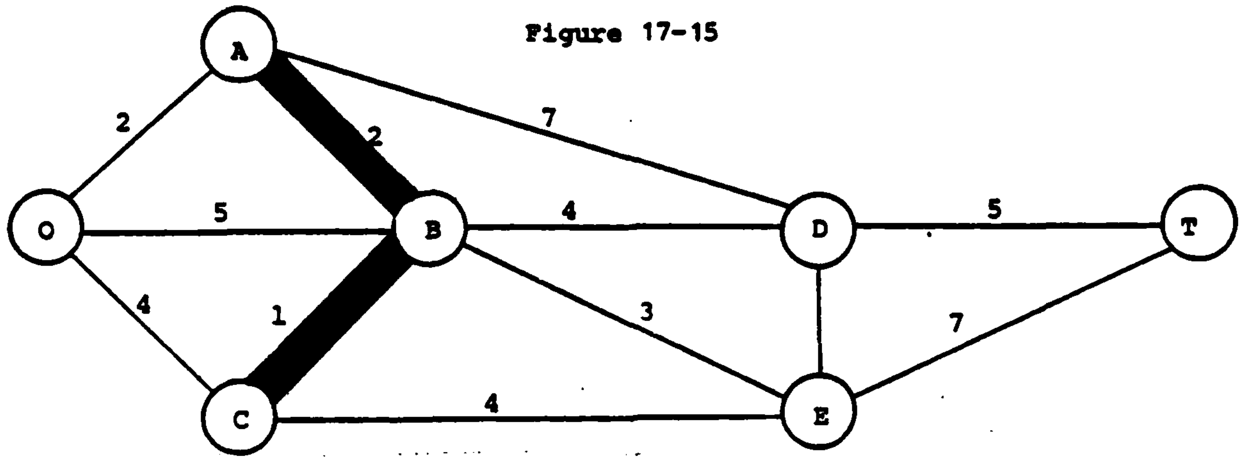
1. Arbitrarily select B as the starting node. Connect B to C as the closest unconnected node.

Figure 17-14



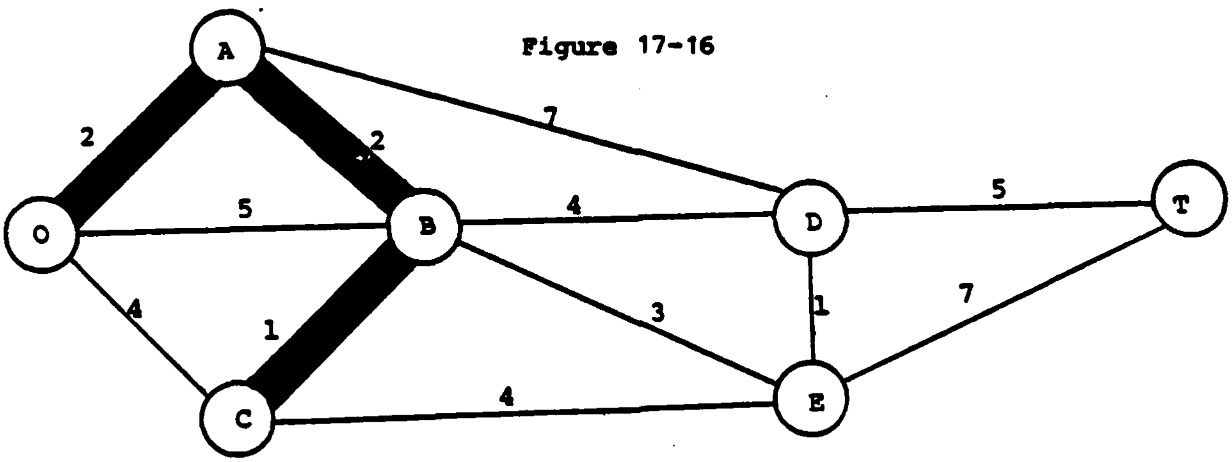
2. A is the unconnected node closest to either node B or C. Connect A to B. Now A, B, and C are connected:

Figure 17-15



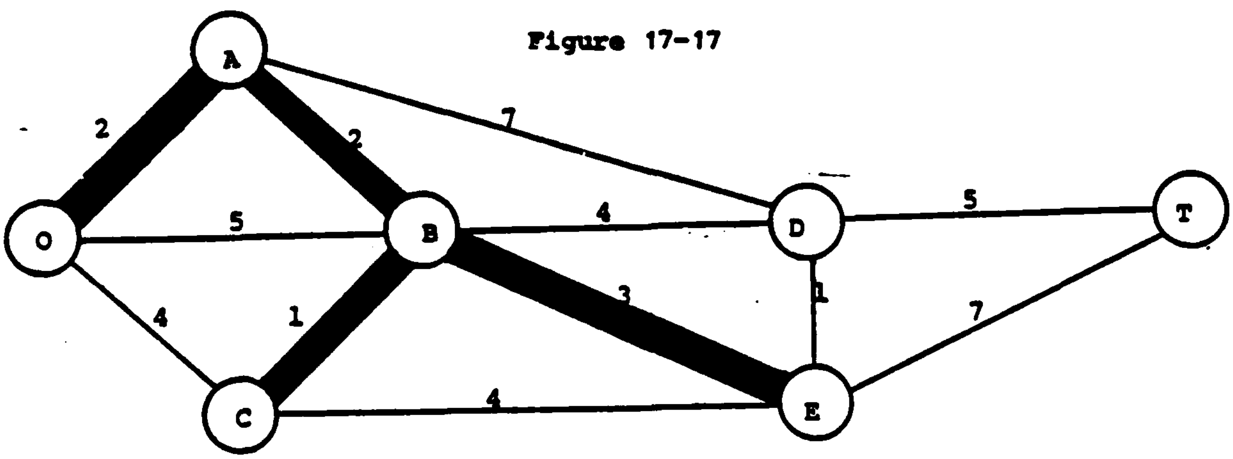
3. 0 is now the unconnected node nearest to A, B, or C. Connect 0 to A. Now 0, A, B, and C are connected:

Figure 17-16



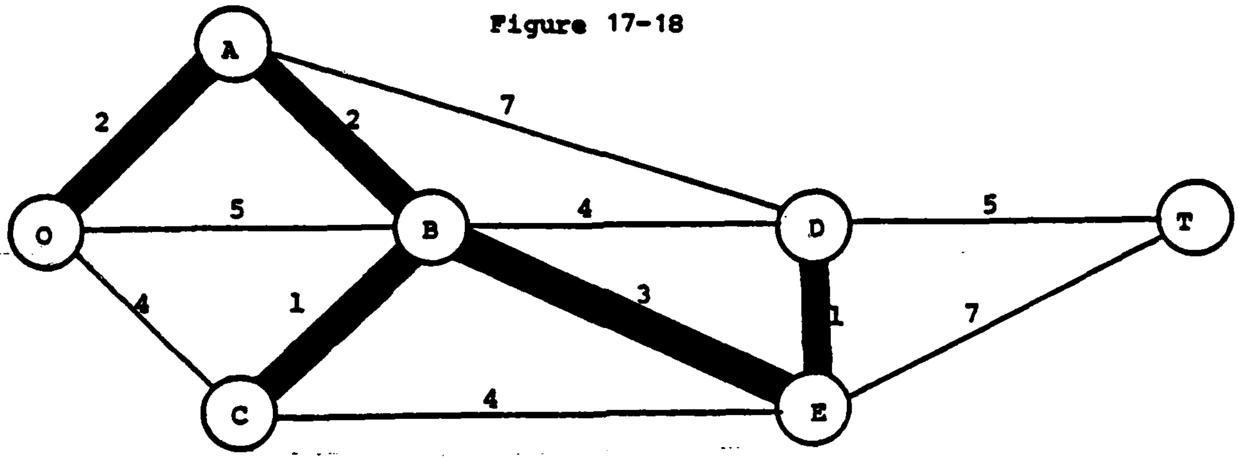
4. E is the unconnected node nearest to O, A, B, or C. Connect E to B. Now nodes O, A, B, C, and E are connected:

Figure 17-17



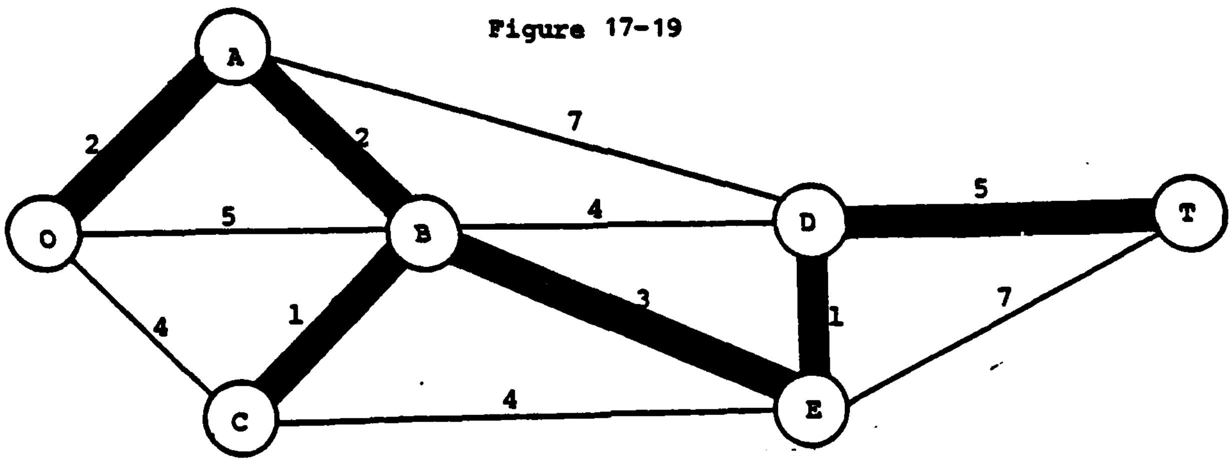
5. D is now the unconnected node nearest to O, A, B, C, or E. Connect D to E. Nodes O, A, B, C, D, and E are now connected:

Figure 17-18



6. Complete the tree by connecting T to D. The completed tree is:

Figure 17-19



The minimal spanning tree value is:

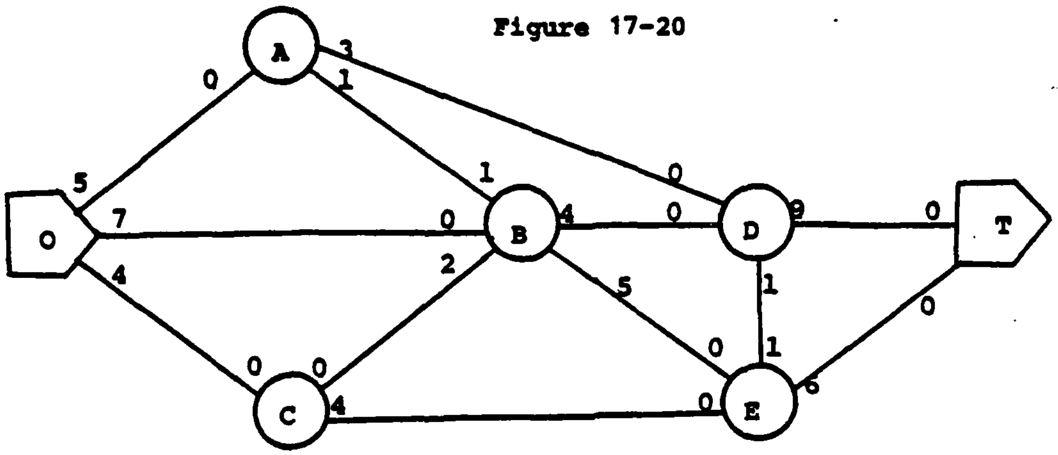
O-A	2
A-B	2
B-C	1
B-E	3
E-D	1
D-T	5
TOTAL	14

THE MAXIMAL FLOW PROBLEM

Problem Description

The maximal flow problem involves determining the maximal amount of total "flow" that can be transmitted through a CAPACITATED NETWORK. For example, consider the following network:

Figure 17-20



The number on a branch (i, j) nearest node (i) represents the capacity of the branch in the direction from (i) to (j). For example, on branch 0-A, there is a flow capacity of 5 units in the direction from 0 to A and a capacity of 0 in the direction from A to 0. The objective is to determine the total amount of flow that can be carried by the network from node 0 to node T.

Stated more formally: Consider a connected network having a single SOURCE and a single SINK. Assume CONSERVATION of FLOW at each node other than the source and sink. Suppose that the rate of flow along branch (i, j) from node i to node j can be any nonnegative quantity not exceeding the specified FLOW CAPACITY (C_{ij}). The objective is to determine the feasible steady-state pattern of flows through the network that maximizes the total flow from the source to the sink.

Solution Techniques

The maximal flow problem can be formulated as a linear programming problem and solved by the simplex method or by any of the available linear programming computer codes, e.g., LINPRO or UHELP. An even more efficient solution procedure is available and is summarized as follows:

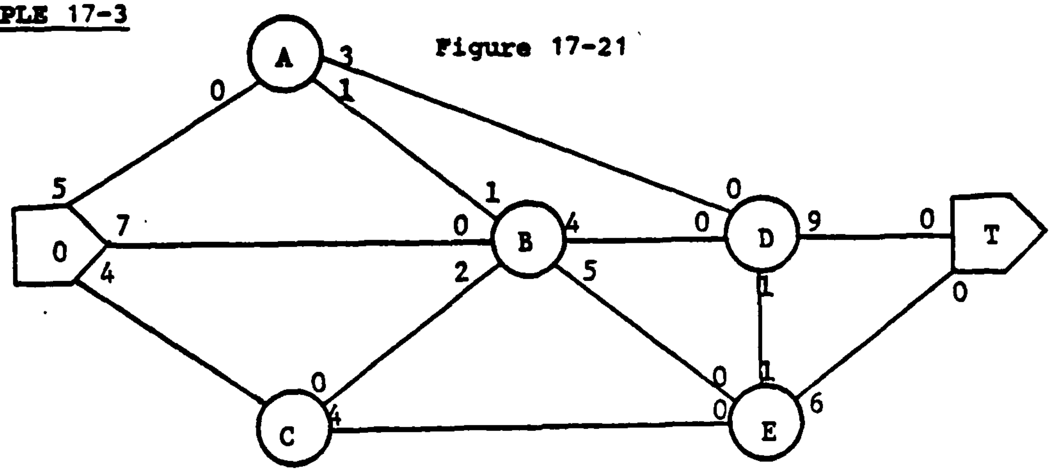
Step 1: Find a path from source to sink with strictly positive flow capacity. (If none exists, the net flows already assigned constitute an optimal flow pattern).

Step 2: Search this path for the branch with the smallest remaining flow capacity (denote this by C*), and increase the flow in this path by C*.

Step 3: Decrease by C* the remaining flow capacity of each branch in the path. Increase by C* the remaining flow capacity in the opposite direction for each branch in the path. Return to Step 1.

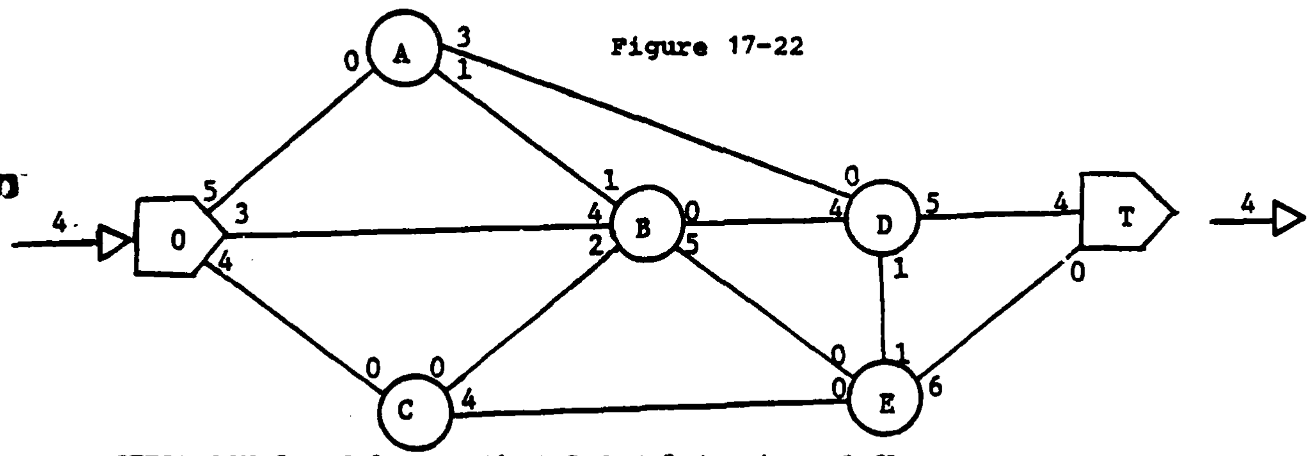
EXAMPLE 17-3

Figure 17-21



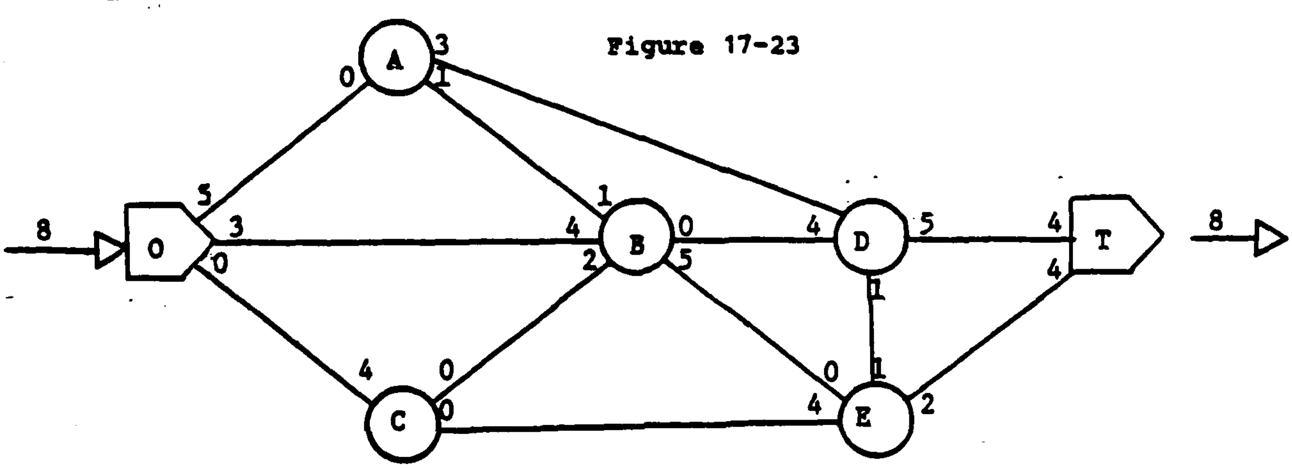
ITERATION 1: Arbitrarily select a path through the network: 0-B-D-T at a flow of C*=4. Adjust the remaining flow capacity:

Figure 17-22



ITERATION 2: Select path 0-C-E-T @ 4 units of flow:

Figure 17-23



ITERATION 3: Select path 0-A-D-T at a flow of C*=3:

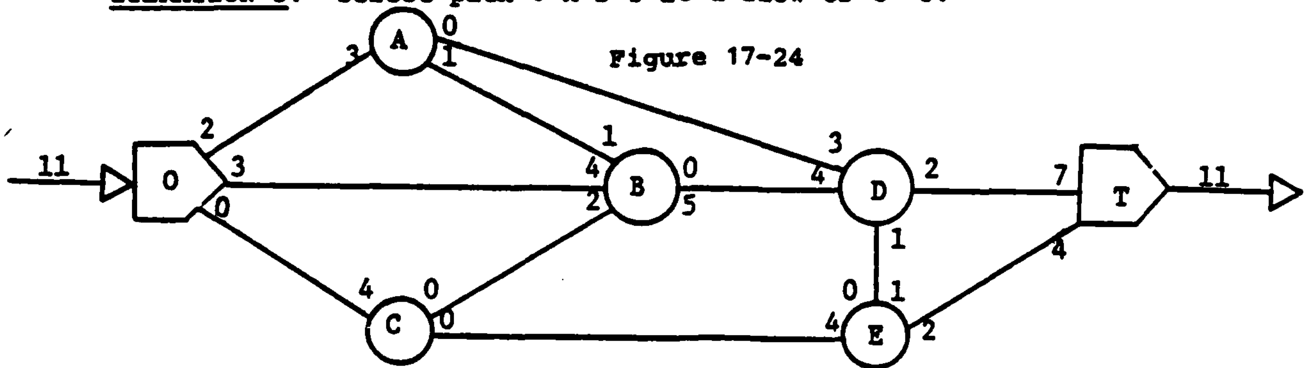


Figure 17-24

ITERATION 4: Select a path 0-A-B-E-T at a flow of C*=1:

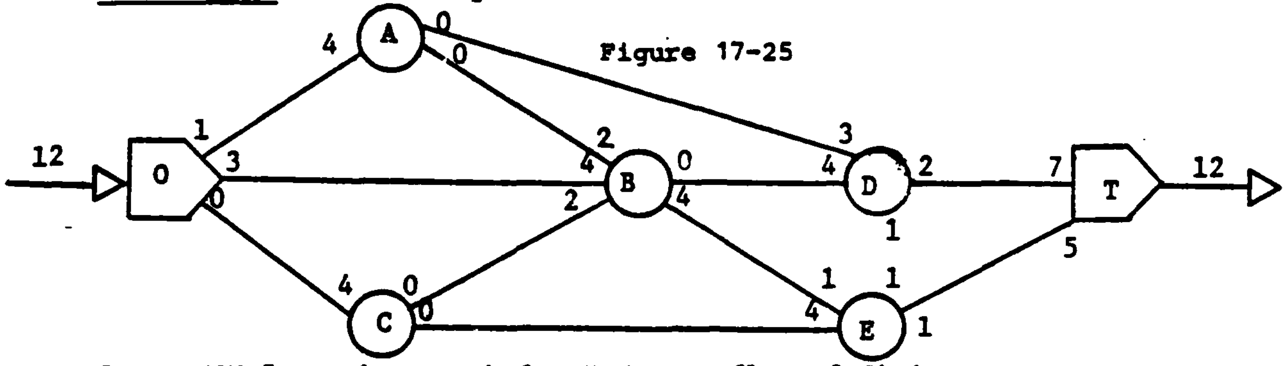


Figure 17-25

ITERATION 5: Select path 0-B-E-T at a flow of C*=1:

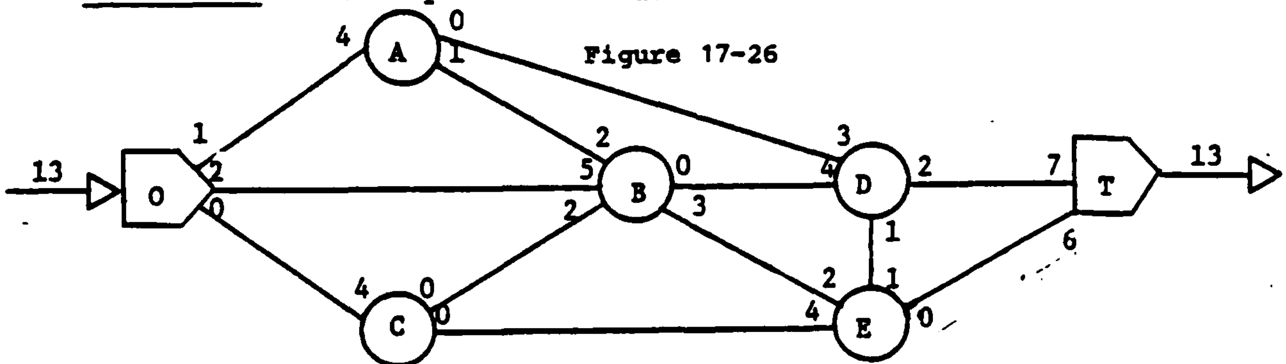


Figure 17-26

ITERATION 6: Select path O-B-E-D-T at a flow of C*=1:

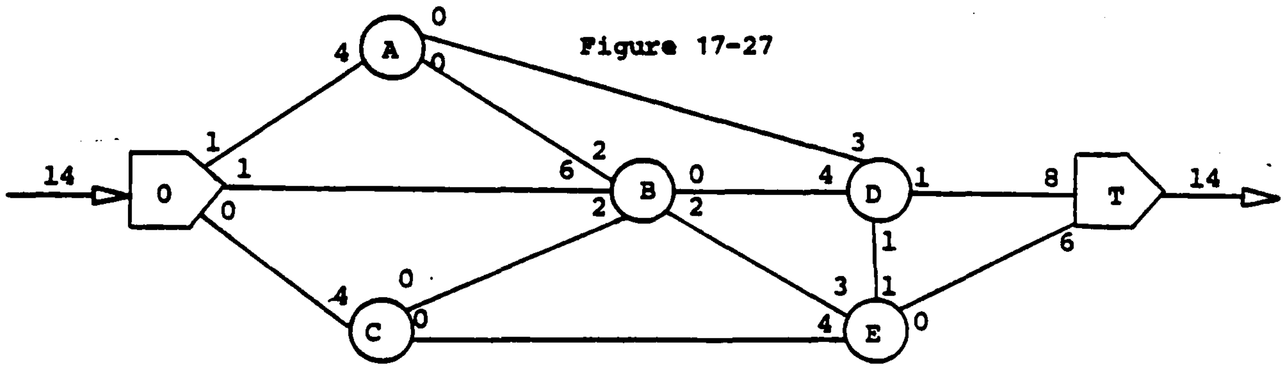


Figure 17-27

No additional improvement can be made. The maximal flow is 14 units.

CLOSED CIRCUIT/TRAVELING SALESMAN PROBLEM

Problem Description

In the closed circuit or traveling salesman problem, the object is to determine the minimal circuit through a network that encounters each node once and only once and returns to the starting node.

As in the case of the minimal path problem, the values associated with each branch may represent time, distance, cost, risk, or other measure of utility.

The problem has become well-known and has been studied for years. This attention is because the problem is easy to state but computationally difficult to solve. In addition, much of the attention given this problem is due to the wide range of problem situations that can be modeled by this formulation:

1. police patrol

2. refuse pick-up

3. preventive maintenance routes

4. staff assistance visits

Obviously, a solution exists for the problem and the minimal circuit could be determined by explicit and/or implicit enumeration. However, considering the combinatorial nature of the problem, the difficulty arises when you realize that in a network of n nodes, there are (n-1)! feasible routes. A network of only 10 nodes, for example, generates 9! or 362,880 routes.

Solution Techniques

A number of solution approaches have been recommended including dynamic programming and 0-1 integer programming.

The integer programming formulations are commonly solved with various "branch-and-bound" algorithms. The technique basically partitions the problem into successively smaller and smaller problems, eliminating routes from further consideration on the basis of implicit enumeration and evaluation. The partitioning process continues until a single optimal route remains.

EXAMPLE 17-4

An inspector wishes to visit each of six project sites and return to his home office. The following table summarizes the distances of the most

direct routes between each pair of project locations:

Table 17-8

TO SITE: (j):

		1	2	3	4	5	6
	1	M	27	43	16	30	26
FROM SITE	2	7	M	16	1	30	25
	3	20	13	M	35	5	0
(i)	4	21	16	25	M	18	18
	5	12	46	27	48	M	5
	6	23	5	5	9	5	M

(Note: The "M" is a very large number introduced to model the condition that the inspector must move from site i to site j - he cannot remain at one place).

In what seqience should the inspector visit the sites to minimize the total distance traveled?

1. The first step in the solution algorithm is to determine the minimum distance in each row and column and to subtract that quantity from each of the other elements in that row or column:

Table 17-9

	1	2	3	4	5	6	ROW CONSTANTS
1	M	11	27	0	14	10	16
2	6	M	15	0	29	24	1
3	20	13	M	35	5	0	0
4	5	0	9	M	2	2	16
5	7	41	22	43	M	0	5
6	18	0	0	4	0	M	5
							43

Table 17-10

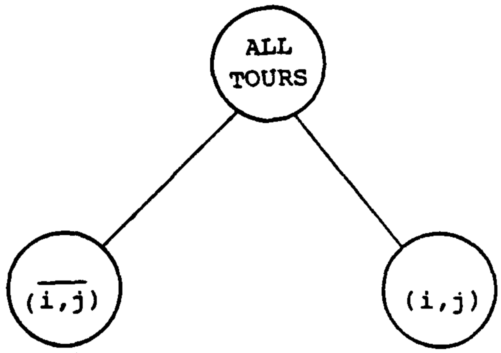
	1	2	3	4	5	6	
1	M	11	27	0	14	10	
2	1	M	15	0	29	24	
3	15	13	M	35	5	0	
4	0	0	9	M	2	2	
5	2	41	22	43	M	0	
6	13	0	0	4	0	M	
Column Constant	5	0	0	0	0	0	= 5

Note that each row and column now contains at least one "0". The distances subtracted from each row and column should be added: 43 + 5 = 48. This forms a "lower" bound on the set of all tours:

ALL TOURS $Z = 48$

2. The next step is to partition this set of all tours into two subsets. This is done by selecting some particular branch (i,j) and using this to partition the set of all tours into a subset of tours containing (i,j) and a complimentary subset of tours which do not contain the branch (i,j):

Figure 17-28



The particular branch selected as the basis for this partitioning is determined by calculating opportunity costs or penalties for each 0-entry in the table. This is done to identify the additional cost or penalty incurred if that branch (i,j) designated by a particular zero is not included in the circuit or tour.

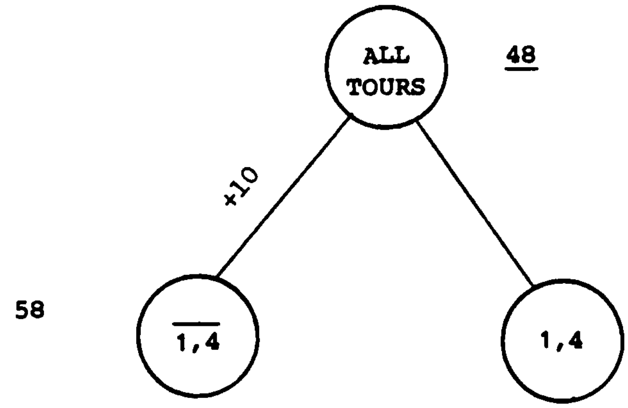
a. The penalty for each 0 entry is calculated by adding together the next smallest values in row (i) and column (j). This value is entered in the cell (i,j) and circled. For example, for the 0 entry in (1,4), the next smallest entry in row (1) is 10. Similarly, the next smallest entry in column (4) is 0. The total penalty for (1,4) is 10+0=10. The penalties for each of the 0 cells are computed in a similar fashion:

Table 17-11

	1	2	3	4	5	6
1	M	11	27	0 10	14	10
2	1	M	15	0 1	29	24
3	15	13	M	35	5	0 5
4	0 1	0 0	9	M	2	2
5	2	41	22	43	M	0 2
6	13	0 0	0 9	4	0 2	M

b. The branch having the largest penalty is selected, in this case (1,4). All of those tours which do not include the branch (1,4) will incur an additional penalty (distance) of 10:

Figure 17-29



3. Since we have examined row (1) and column (4), these can be eliminated from further consideration, and the matrix is collapsed by removing these:

Table 17-12

	1	2	3	5	6
2	1	M	15	29	24
3	15	13	M	5	0
4	M	0	9	2	2
5	2	41	22	M	0
6	13	0	0	0	M

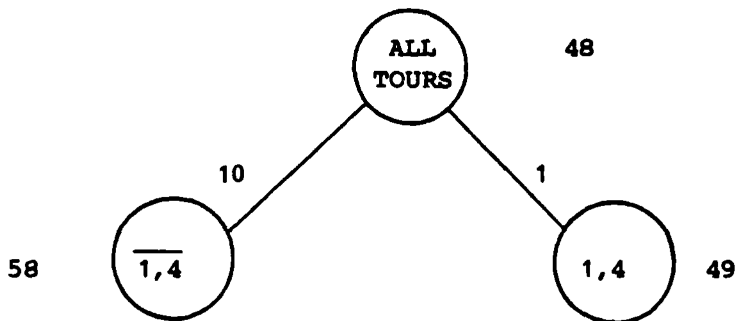
In addition, we need to take some action to preclude "subtours", i.e., selecting a branch that would take us back to a previously connected node before all n(6) nodes have been linked. In this case, we need to insure that we don't select (4,1). This is done by placing an "M" in that position. This completes one iteration of the algorithm.

4. To determine the lower bound on (1,4), we again proceed as in (1) above by "reducing" the matrix, i.e., by subtracting minimum row and column values to create at least one 0 in each row and column:

Table 17-13

	1	2	3	5	6	
2	0	M	14	28	23	1
3	15	13	M	5	0	0
4	M	0	9	2	2	0
5	2	41	22	M	0	0
6	13	0	0	0	M	0
	0	0	0	0	0	1

Figure 17-30



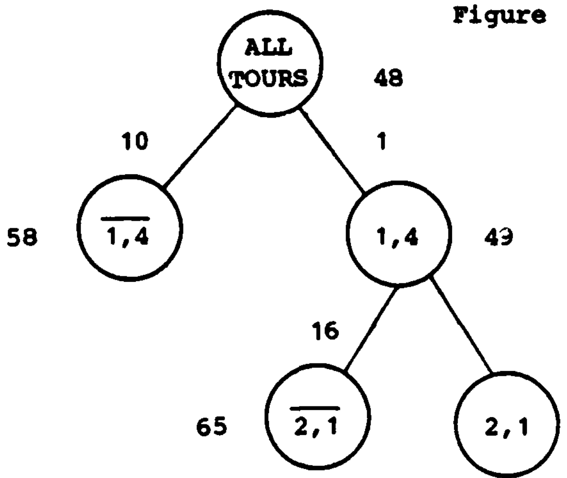
5. To determine the next branch, we again compute the cell penalties for those cells having 0 entries:

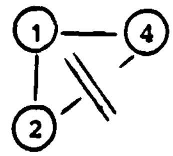
Table 17-14

	1	2	3	5	6
2	16 0	M	14	28	23
3	15	13	M	5	5 0
4	M	2 0	9	2	2
5	2	41	22	M	2 0
6	13	0 0	9 0	2 0	M

Select branch (2,1):

Figure 17-31





6. The matrix is again collapsed by eliminating row (2) and column (1). To prevent subtours, an M is introduced into cell (4,2):

Table 17-15

	2	3	5	6
3	13	M	5	0
4	M	9	2	2
5	41	22	M	0
6	0	0	0	M

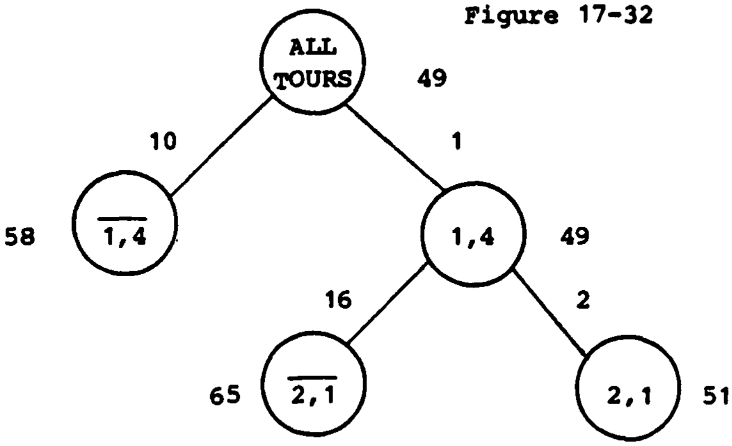
7. We reduce this matrix to determine the lower bound (2,1):

Table 17-16

	2	3	5	6	
3	13	M	5	0	0
4	M	7	0	0	2
5	41	22	M	0	0
6	0	0	0	M	0
	0	0	0	0	2

Therefore, the lower bound is 49 + 2 = 51:

Figure 17-32



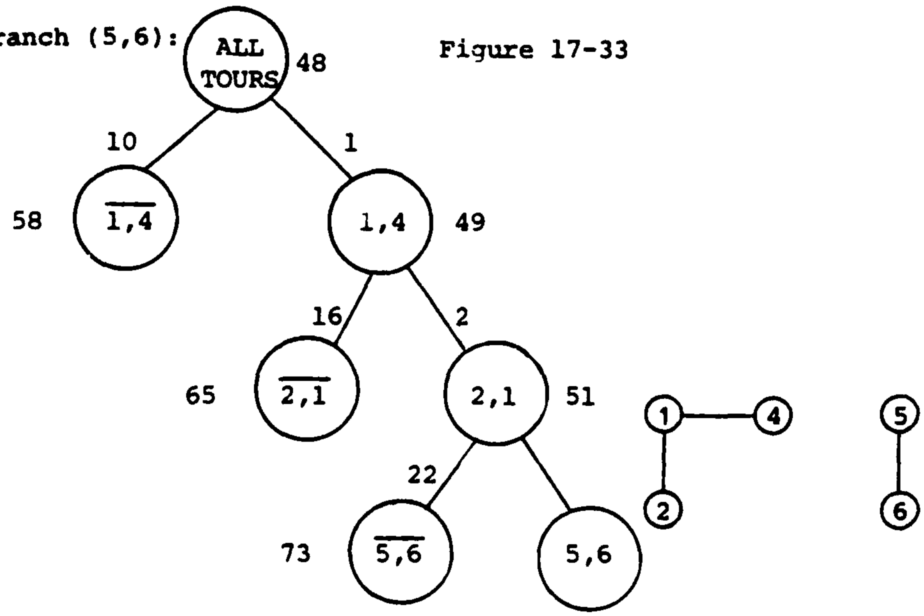
8. Determine the next branch by computing the cell penalties:

Table 17-17

	2	3	5	6
3	13	M	5	0 5
4	M	7	0 0	0 0
5	41	22	M	0 22
6	0 13	0 7	0 0	M

Select branch (5,6):

Figure 17-33



9. Collapse the matrix and introduce an M into (6,5) to prevent subtours:

Table 17-18

	2	3	5
3	13	M	5
4	M	7	0
6	0	0	M

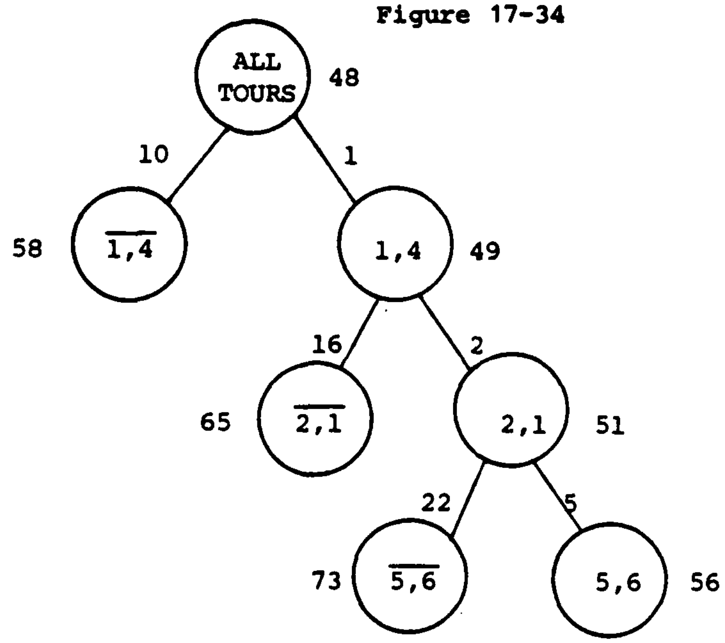
10. Reduce the matrix to determine the lower bound on (5,6):

Table 17-19

	2	3	5	
3	8	M	0	5
4	M	7	0	0
6	0	0	M	0
	0	0	0	<u>5</u>

Therefore, the lower bound is 51 + 5 = 56:

Figure 17-34



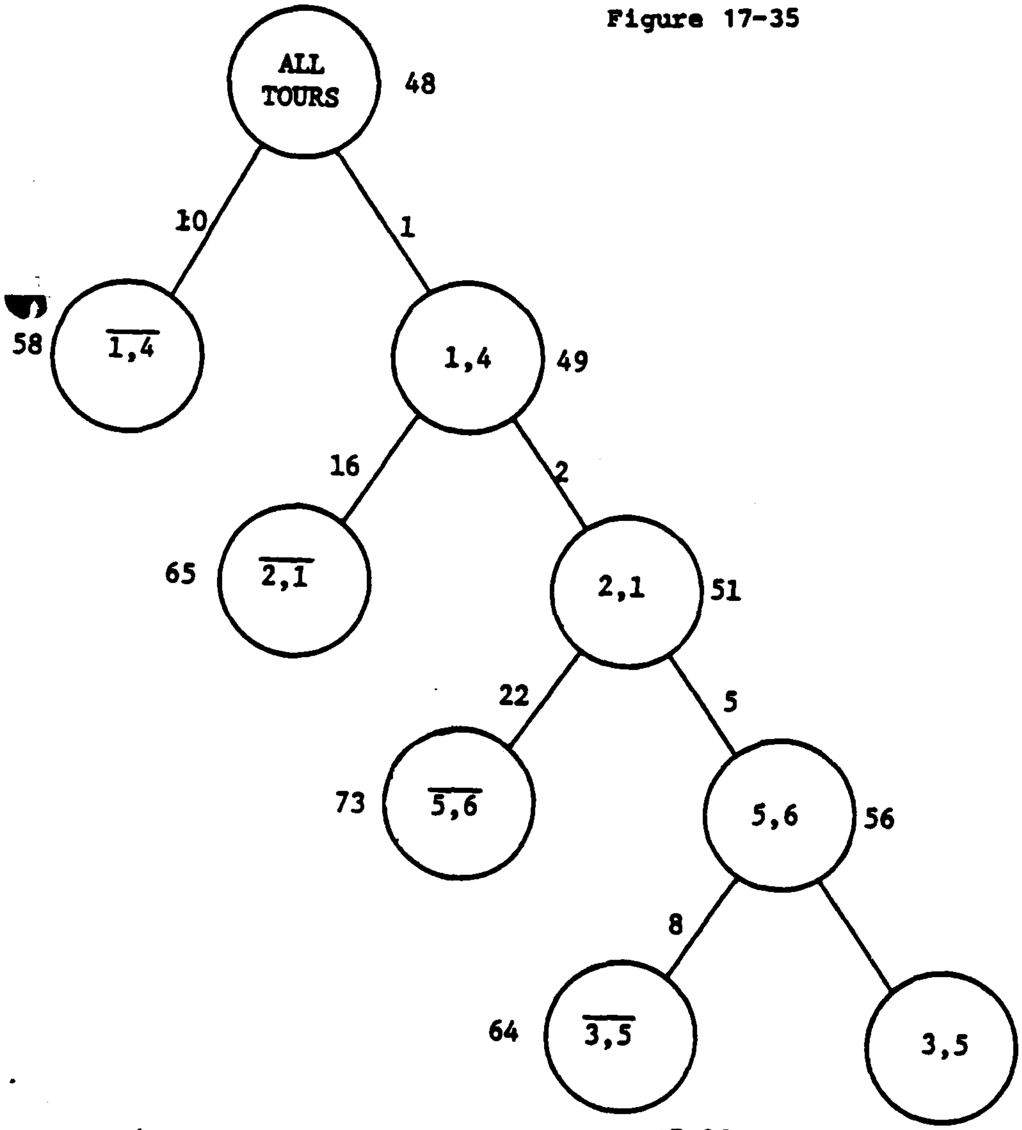
11. Determine the next branch by computing the penalties:

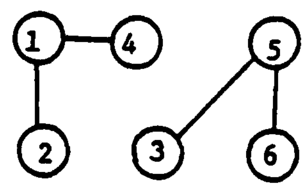
Table 17-20

	2	3	5
3	8	M	0 8
4	M	7	0 7
6	0 8	0 7	M

Note that there is a tie for the cell having the greatest penalty, i.e., (3,5) and (6,2). Arbitrarily select (3,5) as the next branch:

Figure 17-35





12. Collapse the matrix and enter an M in (6,3) to preclude a subtour:

Table 17-21

	2	3
4	M	7
6	0	M

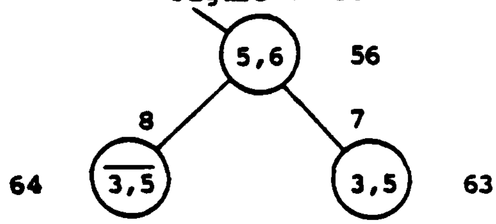
13. Reduce the matrix to determine the lower bound on (3,5):

Table 17-22

	2	3	
4	M	0	7
6	0	M	0
	0	0	7

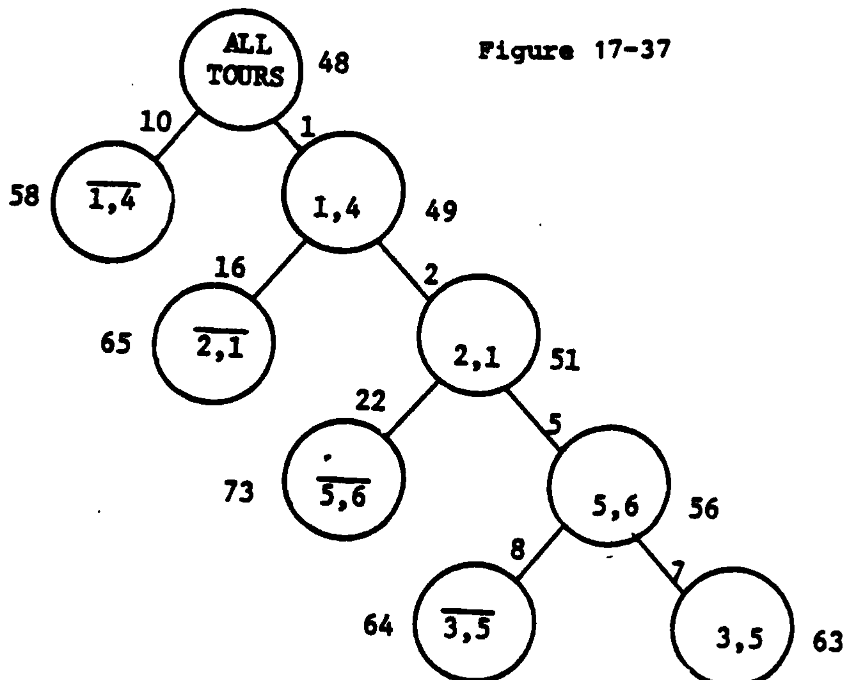
The lower bound is 56 + 7 = 63:

Figure 17-36



14. Now, note that the bound of 63 is no longer minimal. Looking back, we see the lower bound is 58 on ($\overline{1,4}$):

Figure 17-37



There is a possibility (although unlikely) that the optimal tour could be contained in the subset of tours that does not include (1,4). To investigate this possibility, we return to the matrix from which (1,4) was selected as the branch (in this case the original matrix) and preclude the selection of (1,4) by placing an M in that cell:

Table 17-23

	1	2	3	4	5	6
1	M	11	27	M	14	10
2	1	M	15	0	29	24
3	15	13	M	35	5	0
4	0	0	9	M	2	2
5	2	41	22	43	M	0
6	13	0	0	4	0	M

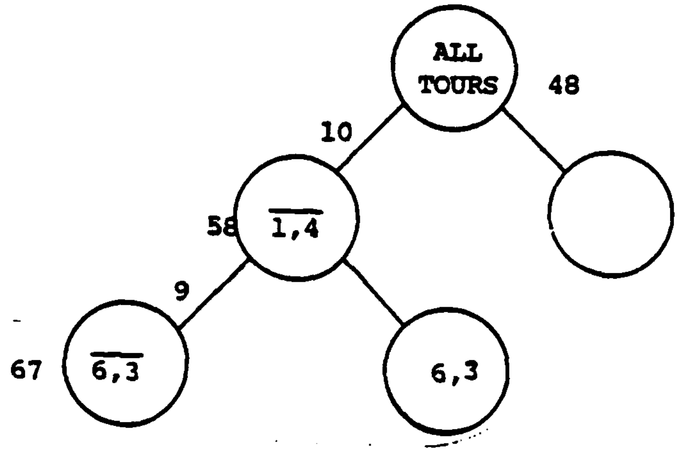
Reducing the matrix and computing penalties:

Table 17-24

	1	2	3	4	5	6	
1	M	1	17	M	4	0 10	10
2	1	M	15	0 5	29	24	0
3	15	13	M	35	5	0 5	0
4	0 1	0 0	9	M	2	2	0
5	2	41	22	43	M	0 2	0
6	13	0 0	0 9	4	0 2	M	10

Select (6,3):

Figure 17-38

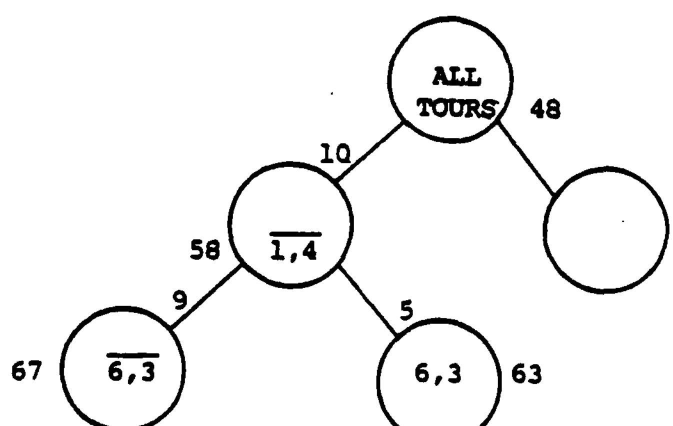


To determine the lower bound on (6,3) we collapse and reduce the matrix and set (3,6) equal to M to preclude a subtour:

Table 17-25

	1	2	4	5	6	
1	M	1	M	4	0	
2	1	M	0	29	24	0
3	10	9	30	0	M	0
4	0	0	M	2	2	5
5	2	41	43	M	0	0
	0	0	0	0	0	5

Figure 17-39



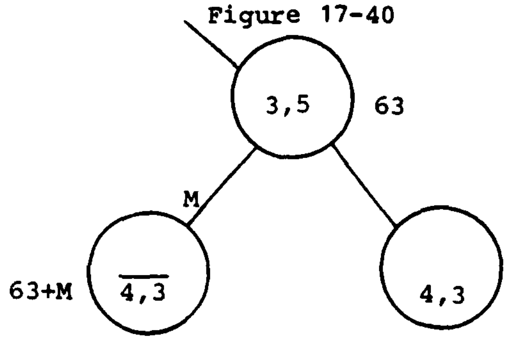
15. At this point we know that the subset of tours which does not include (1,4) but which does include (6,3) must be <u>longer</u> than 63. Since this is not lower than the bound on the subset containing (3,5), we can return to that subset and continue the procedure by determining the next branch:

Table 17-26

	2	3
4	M	0 M
6	0 M	M

Since there again is a tie for the cell having the largest penalty, arbitrarily select (4,3):

Figure 17-40



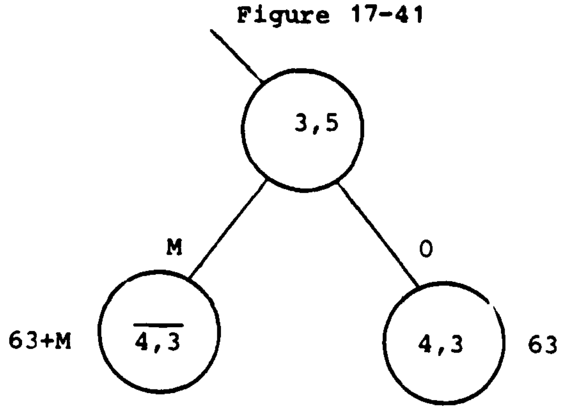
To determine the bound on the subset containing (4,3) we collapse and reduce the matrix:

Table 17-27

	2	
6	0	0

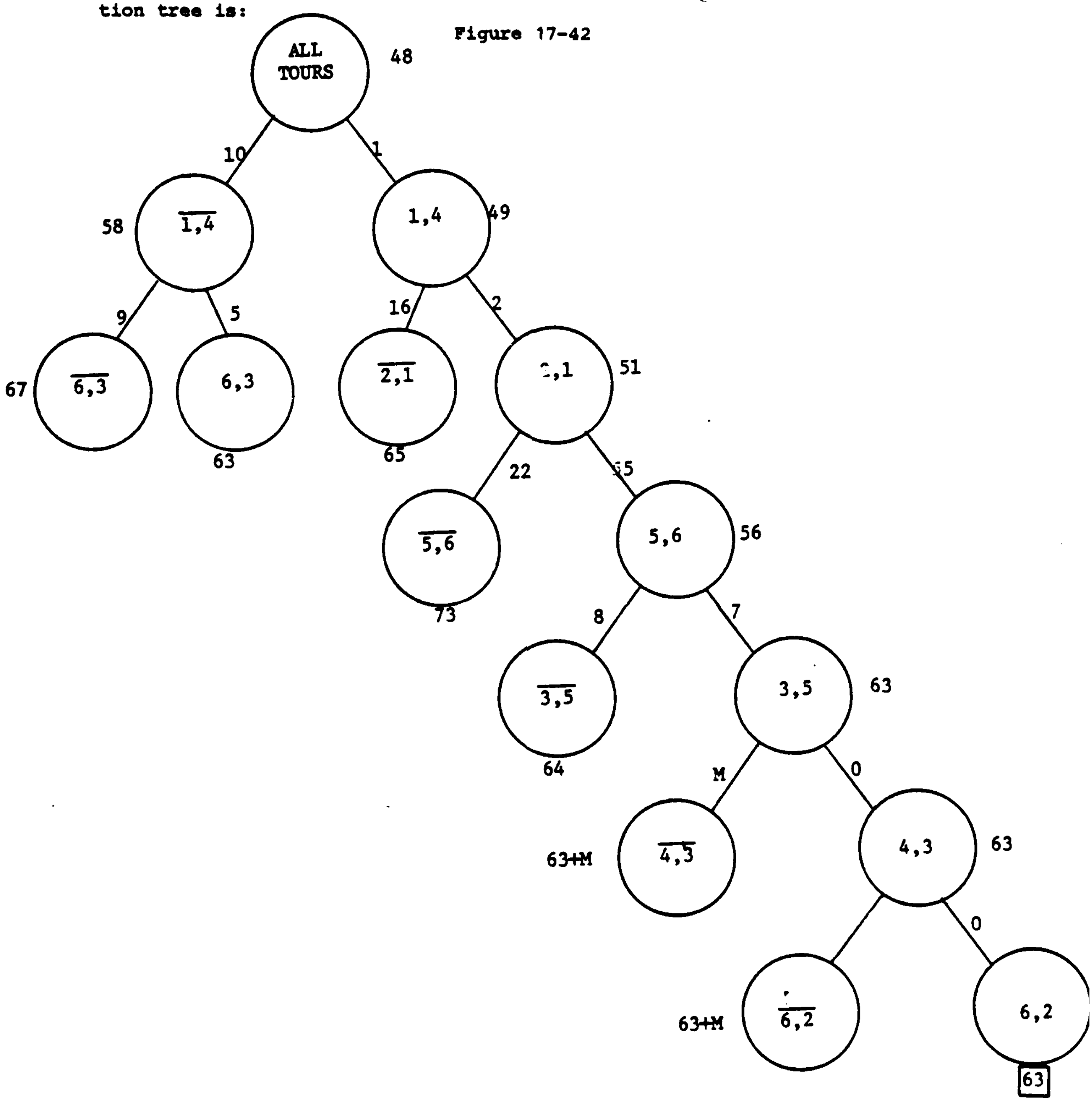
The reduced matrix contains only one cell (6,2) having a 0 entry. Therefore, the bound on this subset is: 63 + 0 = 63

Figure 17-41



The remaining branch (6,2) is automatically determined. The complete solution tree is:

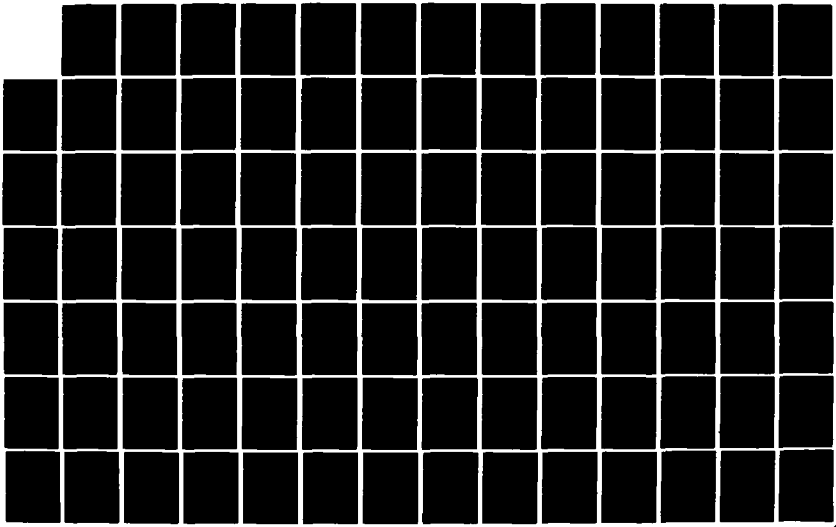
Figure 17-42

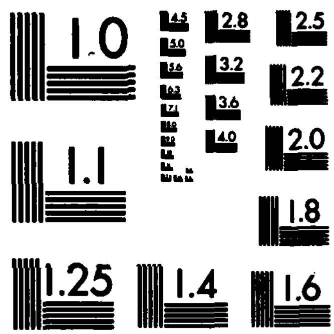


The optimal circuit is: 1 - 4 - 3 - 5 - 6 - 2 - 7

AD-A121 990 QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS J E ENGEL APR 80 AFIT-LS-32 6/7

UNCLASSIFIED F/G 15/5 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

17.2 PROJECT SYSTEM MANAGEMENT

A project is a purposeful activity directed at accomplishing some specific objective. It is usually an activity composed of a number of constituent tasks or "subactivities." These component tasks are usually sequenced or organized by specific precedence relationships. Chase and Aquilano (1973, p. 503) suggest that "A project is a series of related jobs usually directed at some major output and requiring an extensive period of time to perform." In general, a project can be thought of as a system of interrelated discrete task subsystems. Both the system and its constituent subsystems have finite lifetimes. This characteristic is perhaps the most distinguishing feature of a project system.

The term "project management" refers to the management of project systems. More specifically, project management refers to the making (and taking) of decisions concerning the processes of planning, organizing, directing, and controlling project systems. Particular emphasis is given to the planning and controlling processes, with the term planning referring to systems analysis, resource allocation, and scheduling. The control process is generally concerned with identifying operational measures for project system tasks, setting associated objectives, and monitoring progress or performance in meeting these operational objectives.

Project System Analysis

The term "project system analysis" refers to the process of defining, analyzing, and measuring the various elements of a project system, i.e., the component tasks, the relationships between system tasks, and the behavior of the project system as a whole. Operationally, the problem of project system analysis can be distilled down to two fundamental questions:

1. What do we want/need to do? (DECISION)

2. How will we evaluate the results of our efforts? (CONTROL)

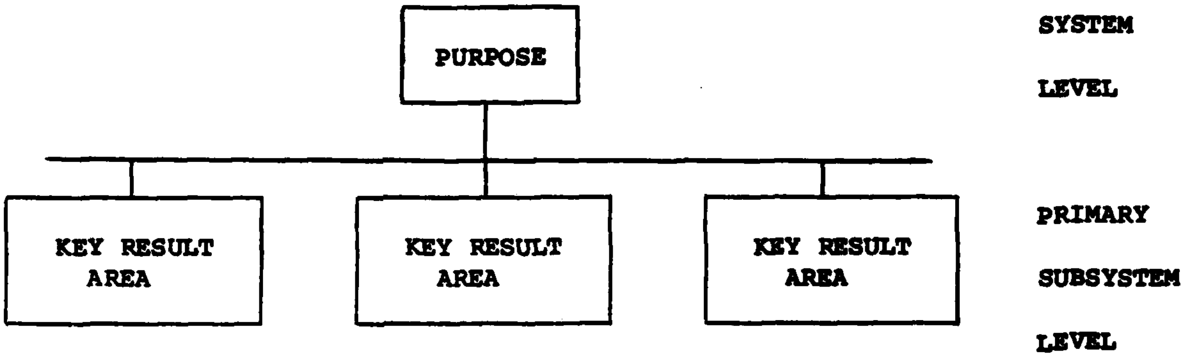
The process of project system analysis can be facilitated by using a conceptual framework to guide our thinking. The framework first focuses on the DECISION question:

PURPOSE: The first step is to explicitly define the purpose of the project system. The purpose is simply a general statement of what we are trying to do or accomplish, e.g., construct a new facility, repair a mechanical system, remove snow from the base, etc.

KEY RESULT AREAS: While the statement of purpose provides our analysis with a general direction, it is usually too broad to serve as a meaningful basis for management action. Consequently, we need to identify the primary subsystems of our total project system. These primary subsystems are commonly referred to as Key Result Areas, because they identify major task areas in which we must insure satisfactory results if the basic

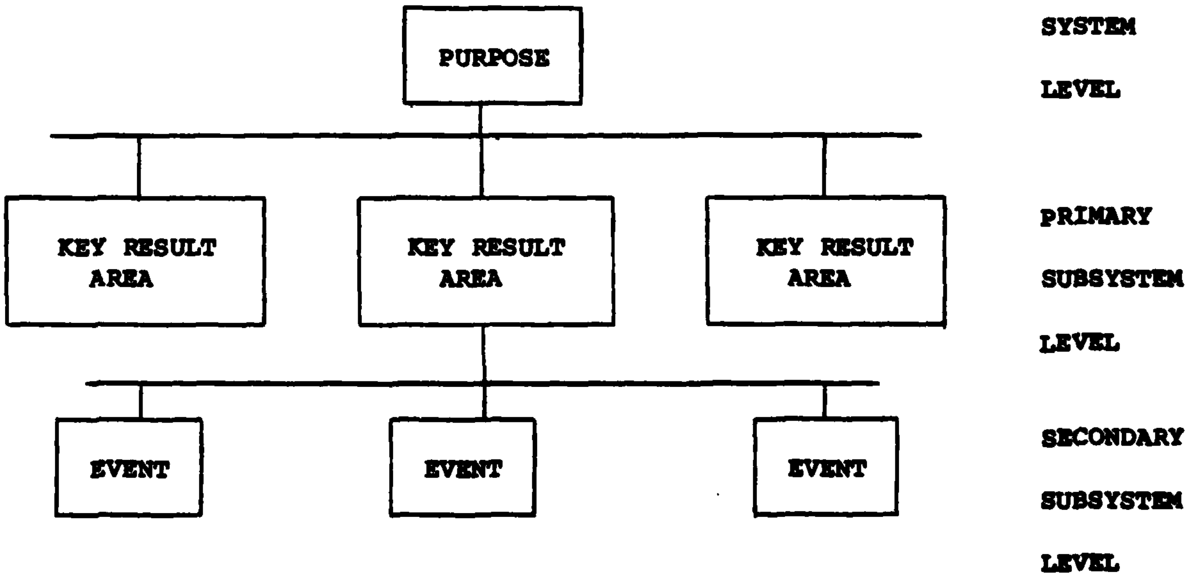
purpose of our project system is to be met. In a facility construction project, for example, Key Result Areas might include structural systems, civil systems, mechanical systems, electrical systems, etc. The relationship between the purpose (system level) and Key Result Areas (primary subsystem level) is a simple hierarchy:

Figure 17-43



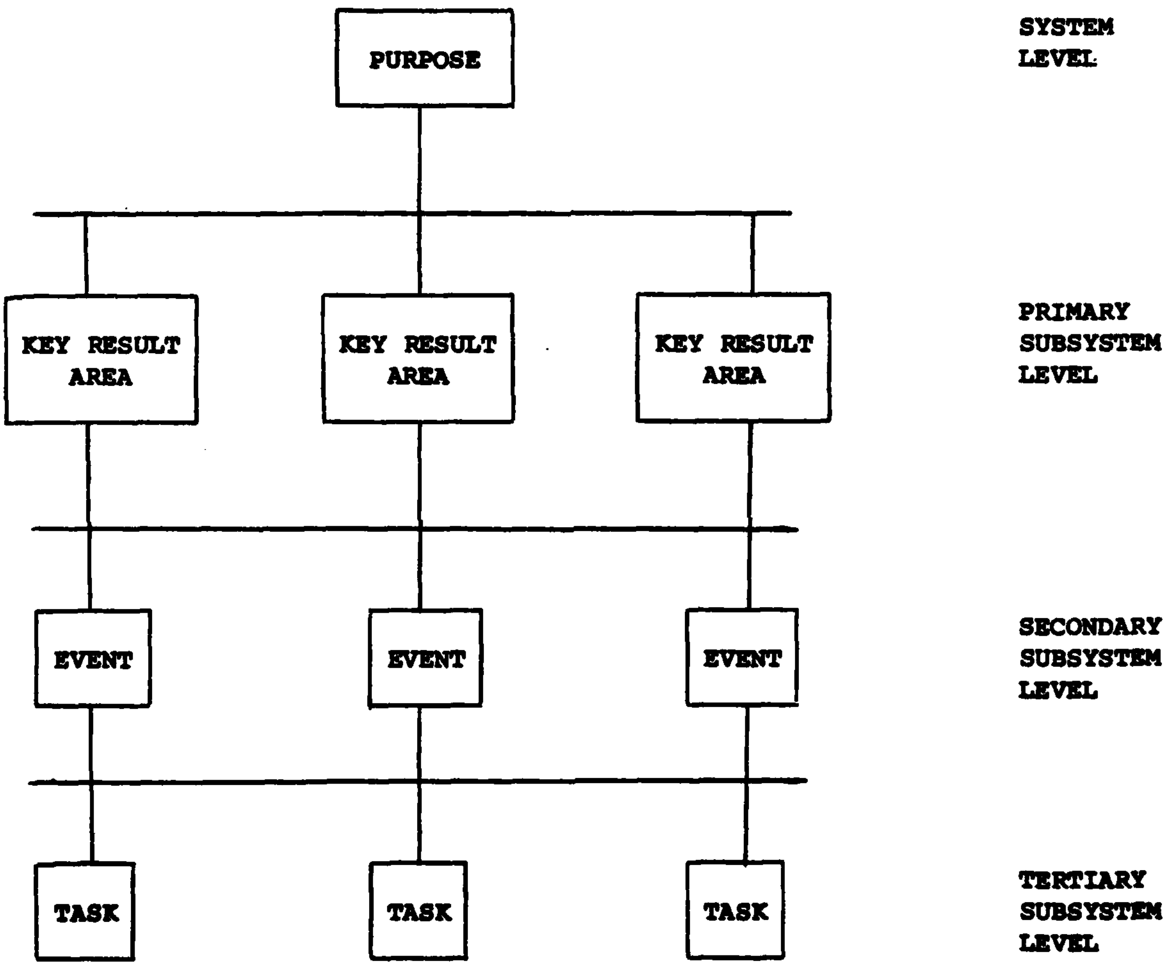
EVENTS: Applying the same logic to each primary subsystem or Key Result Area, we next identify the major tasks or "Events" that define satisfactory accomplishment of the particular Key Result Area. Events correspond to secondary subsystems or major program milestones. For example, considering the mechanical system Key Result Area in the facility construction example, we might define the completion of the plumbing system as one event and completion of the heating and air conditioning system as another event. The relationship between Events, Key Result Areas, and Purpose is illustrated in Figure 17-44.

Figure 17-44



TASKS/ACTIVITIES: Finally, we need to identify those individual component tasks we need to accomplish to complete or realize each (parent) event. Typically, the task is the smallest unit of activity we want to manage, i.e., the greatest level of resolution (tertiary subsystem). The task is the unit to which resources are directly applied, and the level at which performance is measured and controlled. Figure 17-45 illustrates the complete decision structure.

Figure 17-45



The number of levels, or degree of resolution, will vary from project to project and will depend on the degree of control or management visibility required. The more closely we want to control a project, the greater will be the number of levels in the systems analysis structure. Normally, it is not too meaningful or cost effective to go beyond the levels of resolution illustrated here, i.e., beyond the tertiary subsystem level.

Closely tied to these DECISION questions is the complementary question of CONTROL. Essentially, the control question can be restated as "What evidence will we accept that a particular task has been completed or accomplished in a satisfactory manner?" This becomes a question of measurement. More specifically, we need to decide:

1. What operational scales of measurement or "performance criteria" are appropriate for a particular task (e.g., time allowed for task completion in man-days, the compressive strength of concrete in psi, etc.),

2. What our measurement technology and methodology will be, e.g., measurement instrument, measurement frequency, etc., and

3. What level of performance on the specified operational scale or measure is "good enough," i.e., what our operational standard of performance or objective is (e.g., 2.0 man-days for task completion, 2,500 psi compressive strength at seven days for concrete, etc.).

In order to establish appropriate control, and to be consistent, each system task should have <u>at least</u> one and very often many operational measures (with associated performance standards). In many of our projects, we develop extensive plans and specifications to spell out exactly the performance measures to be used and the standards to be met against these measures. For many other projects, our performance criteria, operational measures, and operational objectives are established in our manuals, regulations, and policy letters.

These decisions can be summarized in a DECISION and CONTROL MATRIX for a particular project system. Notice that the matrix illustrated in Figure 17-46 has been embellished to include some additional management information of interest to the project system manager, e.g., the OPR for each task, resources required, current status, etc. Completing such a matrix can give the manager a framework for systematically analyzing the tasks required to complete a project and the associated measures and objectives needed to control it.

While the decision-control matrix facilitates identification of project system task components and desired system behaviors (as expressed in the task performance criteria), it doesn't provide much insight into the relationships (technological, economic, organizational, etc.) between various system components. Determining these relationships depends primarily on the knowledge, experience, and perceptiveness of the project system manager. However, this process can also be facilitated by constructing a matrix of "precedence" relationships.

Figure 17-46

DECISION											PURPOSE
KEY RESULT AREA 2				KEY RESULT AREA 1							
TASK 2.1.2	TASK 2.1.1	TASK 2.2.2	TASK 2.2.1	TASK 1.1.3	TASK 1.1.2	TASK 1.1.1	TASK 1.2.2	TASK 1.2.1			
									TASK O.P.R:		CONTROL
									REQUIRED RESOURCES		
									TECHNIQUE	PERFORMANCE CRITERIA 1	
									FREQUENCY		
									OBJECTIVE OR STANDARD		
									CURRENT STATUS		
									TECHNIQUE	PERFORMANCE CRITERIA 2	
									FREQUENCY		
									OBJECTIVE OR STANDARD		
									CURRENT STATUS		
									TECHNIQUE	PERFORMANCE CRITERIA 3	
									FREQUENCY		
									OBJECTIVE OR STANDARD		
									CURRENT STATUS		

Figure 17-47

	TASK 1.1.1	TASK 1.1.2	TASK 1.1.3	TASK 1.2.1	TASK 1.2.2	TASK 2.1.1	TASK 2.1.2	TASK 2.2.1	TASK 2.2.2
TASK 1.1.1	—		T		T		E		
TASK 1.1.2		—			T				E
TASK 1.1.3			—		T,P				
TASK 1.2.1				—			E		E,T
TASK 1.2.2					—				
TASK 2.1.1						—			T,P
TASK 2.1.2							—	P	
TASK 2.2.1								—	
TASK 2.2.2									—

T = Technological
E = Economic
P = Political

In the matrix illustrated in Figure 17-47, various symbols (e.g., "T", "E", and "P") have been used to denote that a precedence relationship exists between two activities and the nature of that relationship. For example, the technological constraint or relationship indicated by the "T" in cell (1.1.1, 1.1.3) implies that for technological reasons, task 1.1.1 must be completed before task 1.1.3 can be started (e.g., the footing must be completed before a wall can be constructed). The economic relationship indicated in cell (1.2.1, 2.1.2) implies that task 1.2.1 must be completed before task 2.1.2 can be started (e.g., both tasks cannot be accomplished concurrently because they require the same critical piece of equipment or key individual). Obviously, identification of the key relationships between system components is critical. When such relationships are not recognized and identified, we increase the risk of unforeseen delays and/or increases in cost.

PROJECT SYSTEM SCHEDULING AND CONTROL

Once the analysis of the project system has been completed, i.e., we've identified all of the component tasks which need to be controlled and their various interrelationships, we complete the planning process by identifying the resources required and by developing a project system completion schedule. These steps are, of course, closely related to, and integrated with, the system analysis phase of the planning process. Completing the decision-control and precedence matrices described in the previous section generally require the manager to consider the resources required, as well as those available. These resource considerations often influence or determine the relationships between the component tasks of the project system. Very often, subsequent thinking about available and required resources forces us to go back and "fine tune" the relationships developed in our initial project system analysis.

Having completed our system analysis and considered the availability and allocation of required resources, we also need to develop an operating schedule or time-phased plan for completing the project and accomplishing our purpose. Several tools, most notable the bar chart and the activity network, are commonly used in the scheduling process. These techniques are also used extensively in project system control. Consequently, these approaches are commonly discussed under the topics of planning and/or control, depending on the perspective of the respective author.

The Bar Chart

The bar chart (or Gantt chart) is perhaps the most widely used project system scheduling device. A typical bar chart is illustrated in Figure 17-48. This chart can be rather easily constructed from the information compiled in the project system analysis phase and summarized in the decision-control and precedence matrices illustrated in Figures 17-46 and 17-47. The typical bar chart is advantageous in that it is relatively easy to read and highlights activities requiring management attention. In the example of Figure 17-48, the status of day three indicates that Task 1.2.1 is in-progress

and ahead of schedule; Task 1.1.1 has been completed on schedule; Task 1.1.2 is in-progress, but behind schedule; and Task 2.1.1 is in-progress and on-schedule. This basic information can be embellished by including other useful information, e.g., percent complete (for each task), color coding of critical activities, etc. On the other hand, the bar chart has the disadvantage that it does not (normally) clearly show all of the task interrelationships. For example, inspection of the chart in Figure 17-48 suggests that Task 2.2.2 must also be preceeded by Tasks 1.2.1 and 1.1.2. This limitation can be offset to some extent by an additional bookkeeping convention or notation to reflect precedent activities (see, for example, the bar for Task 2.2.2). However, this procedure tends to be somewhat burdensome for large projects.

The Project System Network

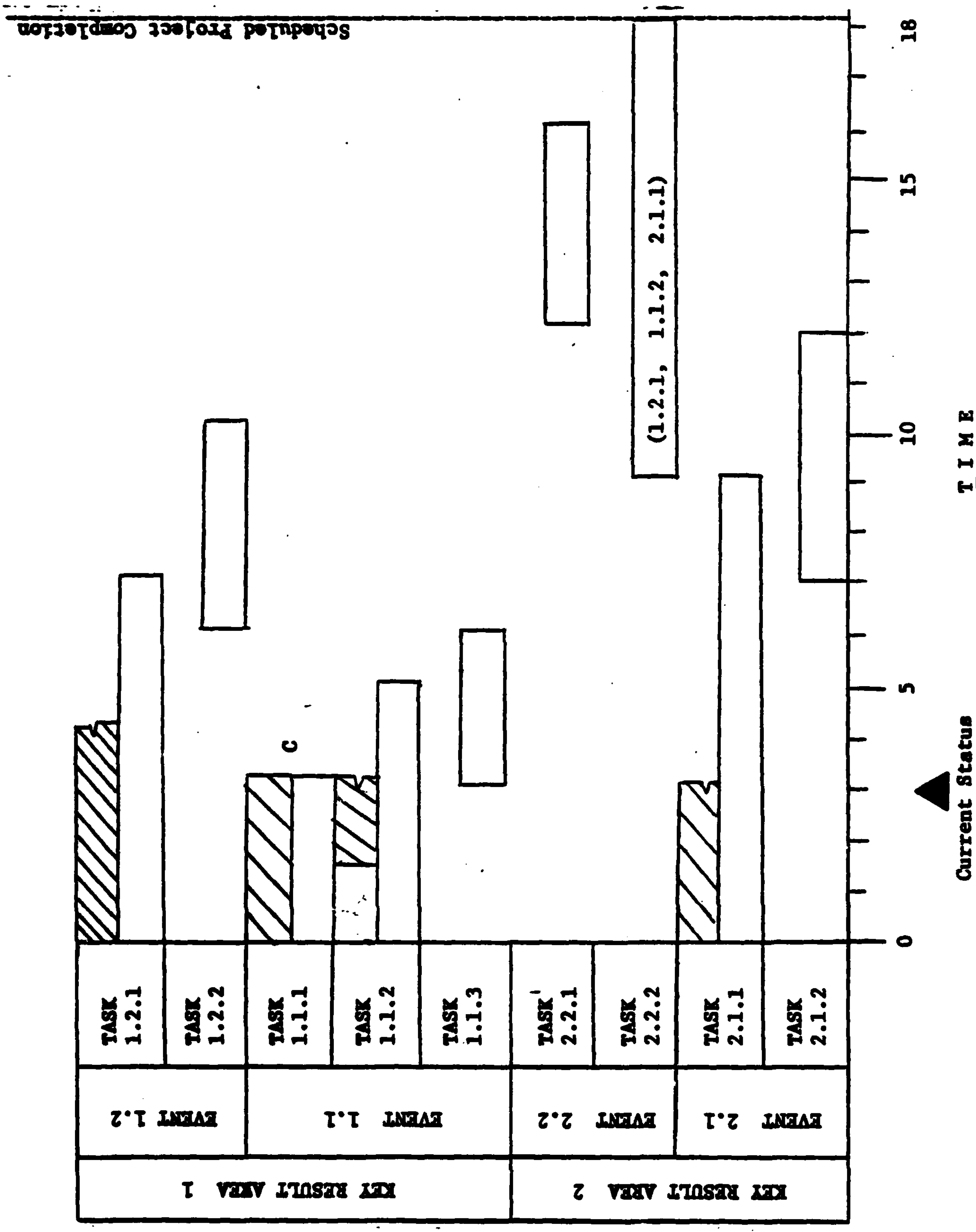
Perhaps a more powerful scheduling and control technique is the project system network or activity network in which the project system is modeled as a network or graph in which the nodes represent events and incident arcs represent activities which must be completed before the event can be said to be completed. A typical project system network is illustrated in Figure 17-49. The project system network has the advantage of more clearly illustrating the interrelationships between various system components. As such, it is particularly useful to the manager in assessing the impact of delays in the completion of particular tasks or conversely, in taking advantage of opportunities created by the early completion of various system tasks. The Program Evaluation and Review Technique (PERT) and the Critical Path Method (CPM) are two commonly used project system management techniques based on the activity network.

The project system network, as normally constructed, is not drawn to a time scale as is the bar chart, posing some limitation is using it for control purposes, i.e., for displaying current status. It is possible, however, to construct a time-scaled network diagram, as shown in Figure 17-50.

Organizing and Directing

In discussing project system management, there seems to be a tendency to concentrate on the planning and control processes, while giving relatively less emphasis to the organizing and directing processes. This happens for a number of reasons. Very often, the project manager doesn't have the authority to make significant changes to the organization structure or to make desired assignments of authority and responsibility to particular individuals. Consequently, there is a tendency to treat the organization as a "given" and modify the plan to accommodate this constraint. Even if so contrained, however, the project manager should at least consider how the organization can be modified, either directly or indirectly, to facilitate project system accomplishment. While the organizing process tends, in general, to follow the planning process, these management activities are obviously interdependent. In developing the system relationships in the analysis phase of the planning process, existing (and desired) organizational relationships must be considered.

Figure 17-48



Scheduled Project Completion
KEY RESULT AREA 1
KEY RESULT AREA 2
EVENT 1.2
EVENT 1.1
EVENT 2.2
EVENT 2.1
TASK 1.2.1
TASK 1.2.2
TASK 1.1.1
TASK 1.1.2
TASK 1.1.3
TASK 2.2.1
TASK 2.2.2
TASK 2.1.1
TASK 2.1.2
C
(1.2.1, 1.1.2, 2.1.1)
0
5
10
15
18
Current Status
TIME

Figure 17-49

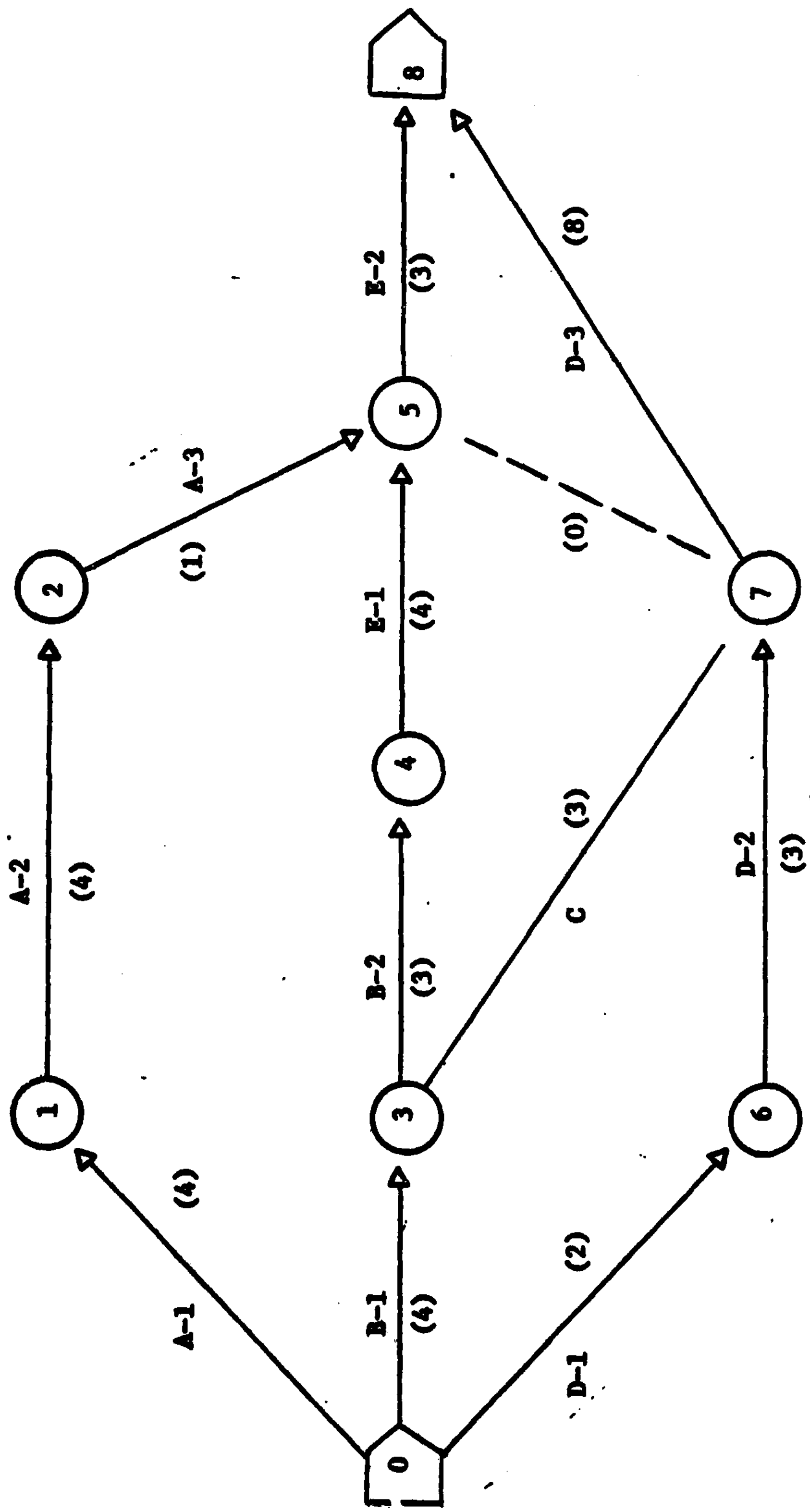
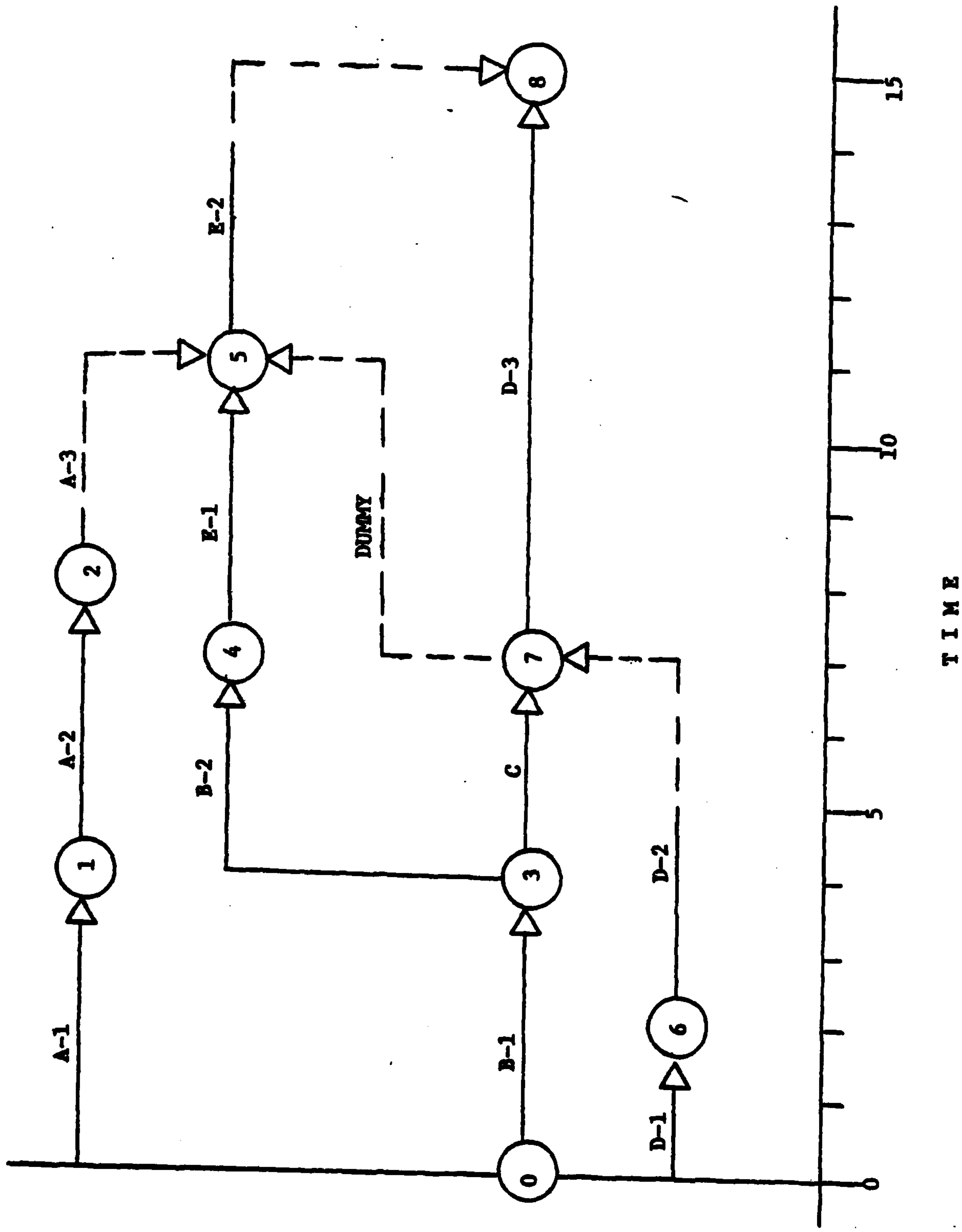


Figure 17-50

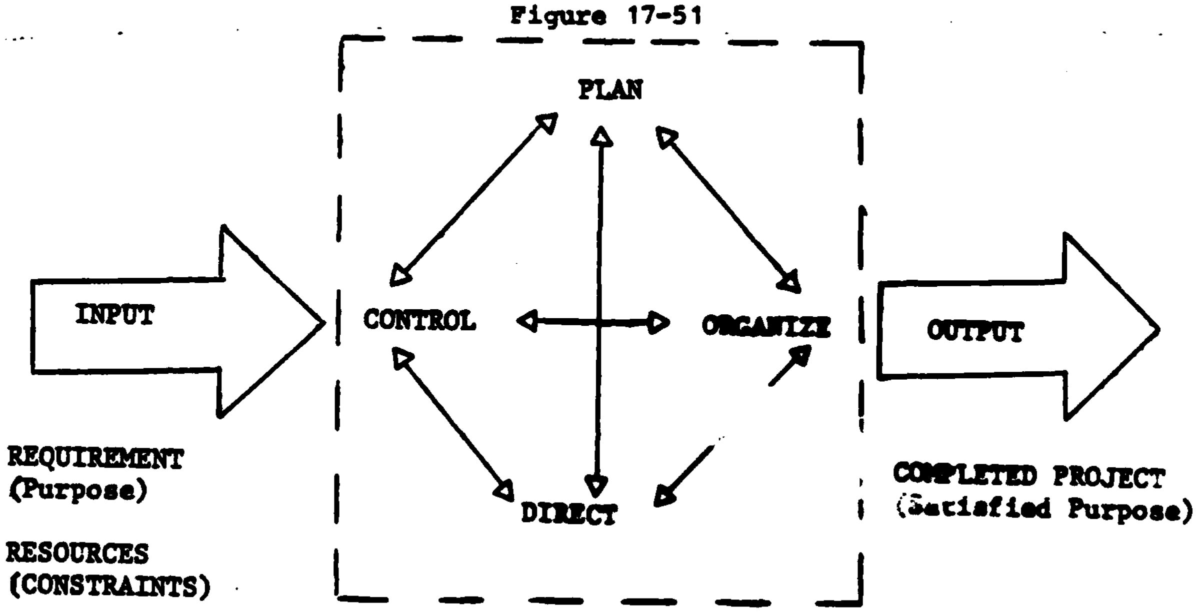


Similarly, success in implementing the project system plan, through the existing or modified organization, depends on how well the project manager communicates the purpose and the details of the plan to those individuals responsible for accomplishing the various component tasks in the project system. The directing/communicating process of management is particularly critical to the success of those projects in which close coordination is required. Typically, as the degree or amount of component interaction, i.e., system complexity increases, the importance of effective direction and communication also increases. In a sense, the directing process is particularly important in that it is through this management process that the manager "breathes life" into the system.

An additional related point should be emphasized concerning effective project management. All project systems depend directly or indirectly on the actions of the people concerned with the project. Because of the unique attributes and characteristics of human resources, the project manager needs to give serious consideration to motivational factors and other organizational behavior consequences associated with project system development and implementation. Very often the difference between success or failure on a project depends to a great extent on the manager's ability to involve and motivate the people involved. In fact, quite typically, these organizational behavior considerations generate the most difficult problems for the project manager as well as the greatest oportunities for success.

The general process of project system management is illustrated in Figure 17-51. This figure attempts to reflect the dynamic input-output nature of the project system as well as the interrelated nature of the various processes involved in management of the system. It should also suggest, at least indirectly, that effective project management is a very complex process, drawing on technical, organizational, and behavioral management skills.

Figure 17-51

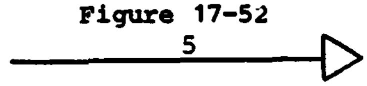


PROJECT SYSTEM NETWORKS

Basic Definitions and Conventions

1. An ACTIVITY is any portion of a project which consumes time or resources and has a definable beginning and ending. Activities are also referred to as tasks or jobs. Activities are graphically represented by directed area (arrows), usually having the associated time estimate (and other pertinent information) written along the arc:

Figure 17-52

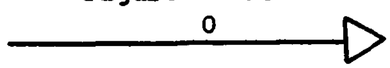


2. A directed arc that does not represent the consumption of time or resources, but indicates a precedence relationship or constraint is called a DUMMY ACTIVITY. It is usually represented by a dashed-line arrow:

Figure 17-53

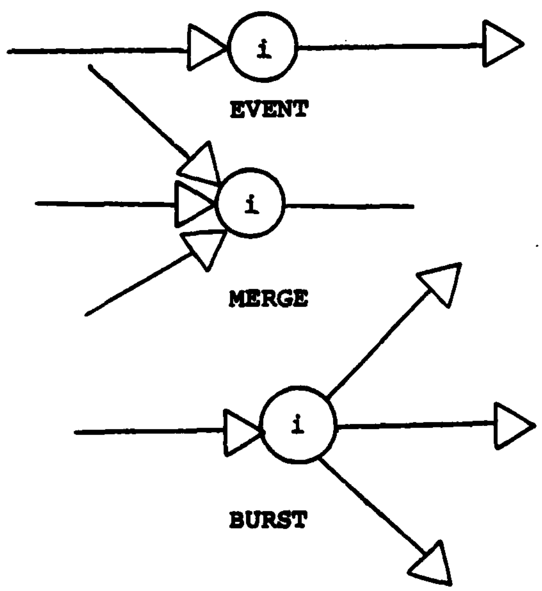
A dummy activity can also be represented by a solid-line directed arc with an associated zero time estimate:

Figure 17-54



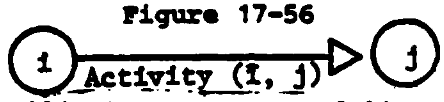
3. The end points of an activity arc represent EVENTS. Theoretically, an event is an instantaneous point in time. Graphically, events are represented by nodes or circles. If an event represents the joint completion of more than one activity, it is called a MERGE EVENT. If an event represents the joint initiation or more than one activity, it is called a BURST EVENT:

Figure 17-55



4. An activity is usually identified by numbers assigned to its beginning and ending nodes:

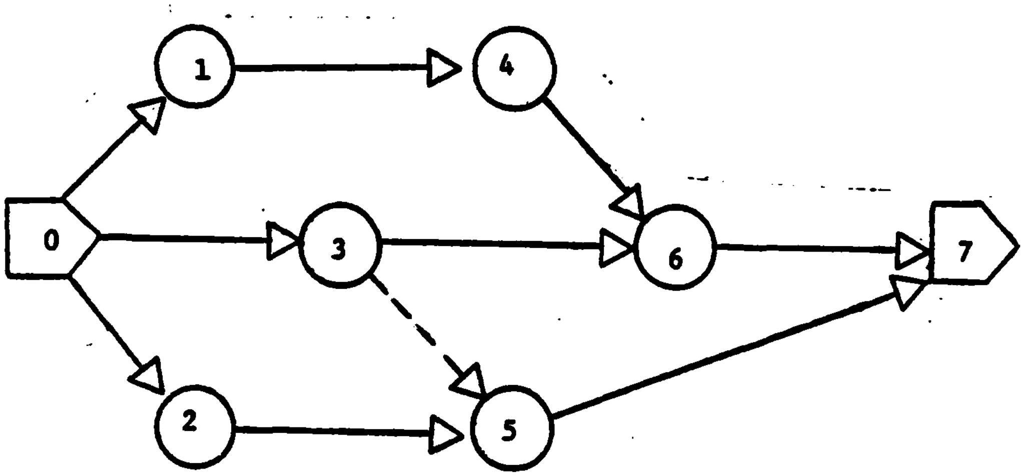
Figure 17-56



By convention, and to facilitate computer modeling, the number label assigned to node (i) is always <u>smaller</u> than the number label assigned to node (j).

5. An <u>ACTIVITY NETWORK</u> is a graphical representation of a project plan, showing the project components (tasks or activities) and the interrelationships (dependencies) between the various activities:

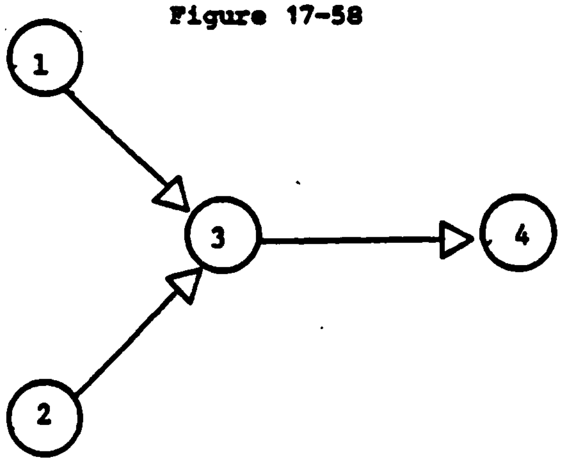
Figure 17-57



<u>Basic Network Rules</u>

Rule 1. Before an activity may begin, all activities preceeding it must be completed. In the following network segment, both activities (1, 3) <u>and</u> (2, 3) must be completed <u>before</u> (3, 4) can be started:

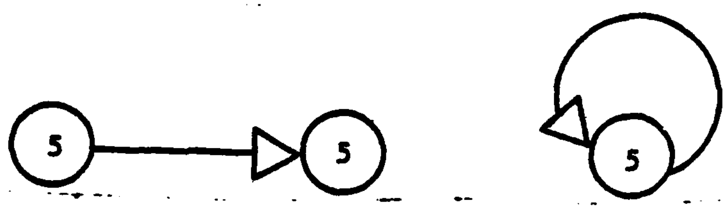
Figure 17-58



Rule 2. Arrows (directed arcs) represent logical precedence only. Normally, neither the length of the arc nor its "compass" direction on the graph have any significance (although some networks are time-scaled to facilitate display of control information).

Rule 3. Event numbers must not be duplicated in a network (i.e., no loops are permitted):

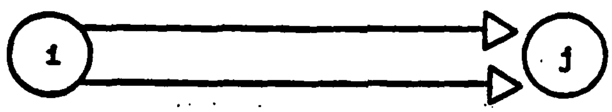
Figure 17-59



NOT PERMITTED

Rule 4. Any two events may be directly connected by no more than one activity:

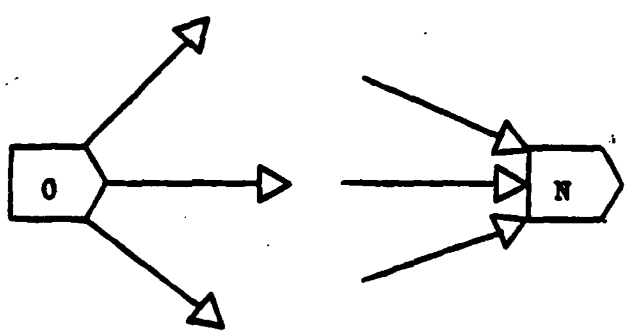
Figure 17-60



NOT PERMITTED

Rule 5. Networks may have only one initial event (i.e., ORIGIN or SOURCE) and only one terminal event (DESTINATION or SINK). It is often desirable to represent the origin and destination nodes by distinctive symbols:

Figure 17-61



ORIGIN

(SOURCE)

DESTINATION

(SINK)

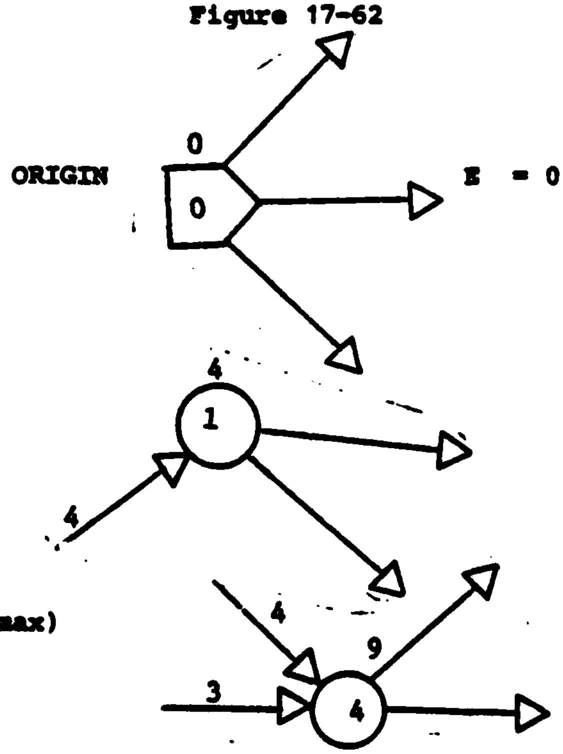
Basic Scheduling Concepts

Once the project system has been analyzed to determine the component activities and their interrelationships, we can construct the project network. The next step (usually done concurrently with network development) is to estimate the time required to complete each activity. The manner in which these time estimates are determined is one of the essential differences between PERT and CPM. Briefly, PERT incorporates probability in the estimate of activity times, i.e., it considers the estimated completion time to be some random variable, while CPM employs deterministic estimates. to introduce some basic scheduling concepts, we will first adopt the deterministic assumption. To facilitate this discussion, consider the project system network of Figure 17-63. The number associated with each activity arrow is the deterministic time estimate. Suppose, as the project manager, you are interested in determining the total project completion time, i.e., the earliest time all activities in the network can be completed. This is equivalent to finding the EARLIEST TIME for each event. Since an event cannot occur until all of the incident activities terminating at the event have been completed, the earliest time of event completion is given by:

$$E_j = \max_i (E_i + t_{ij})$$

where t_{ij} is the estimated activity completion time for (i, j). Normally, E_j is indicated OVER the node to which it applies:

Figure 17-62



Node 1

$E_1 = E_0 + t_{0,1}$
$= 0 + 4 = \underline{4}$

As another example:

Node 4

$E_4 = E_1 + t_{1,4}$
$= 4 + 5 = \underline{9}$ (max)

$E_4 = E_2 + t_{2,4}$
$= 4 + 3 = 7$

When the terminal node of the network has been labeled with its earliest time, the earliest time of project completion has been determined.

Figure 17-63

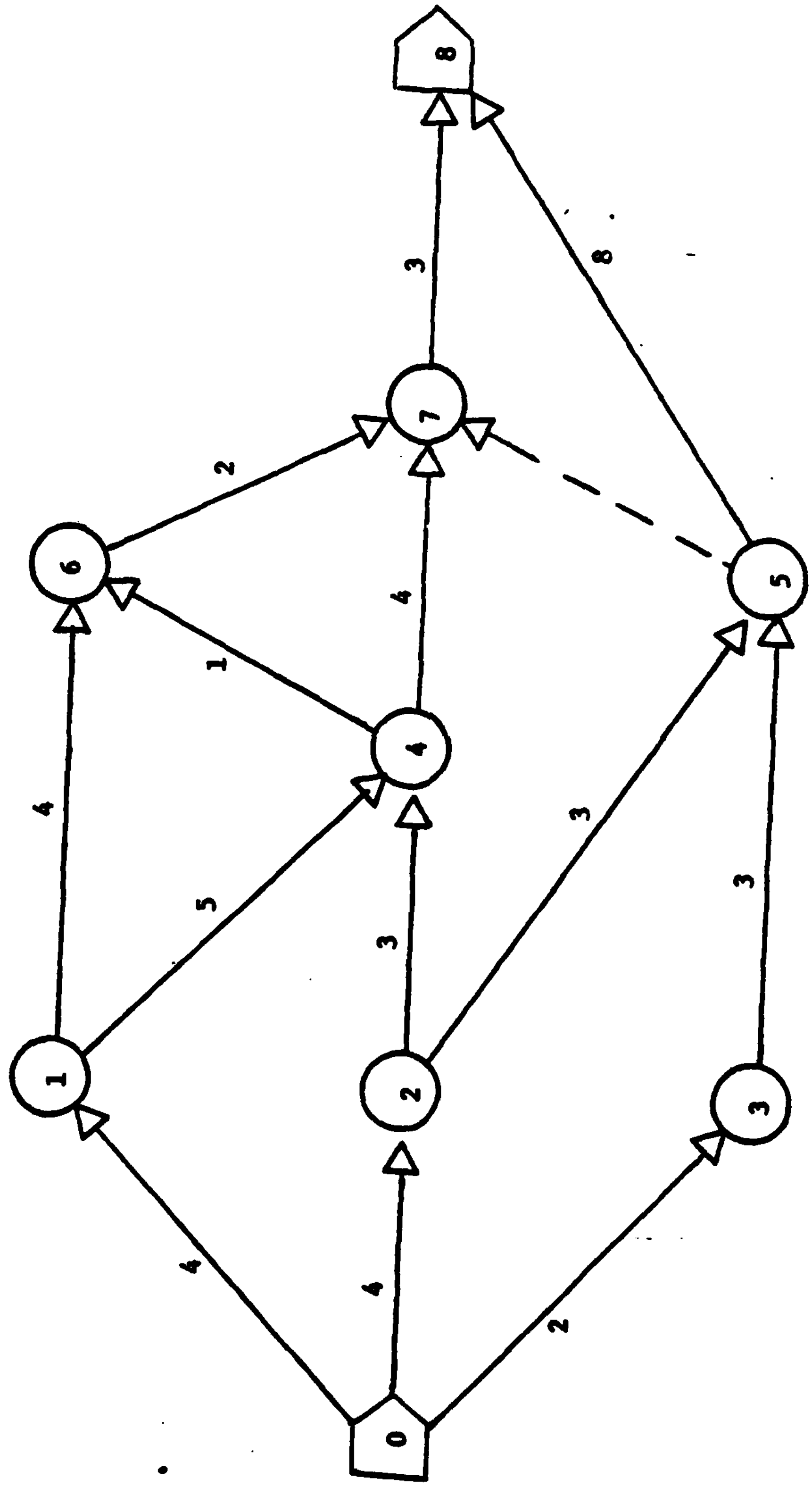
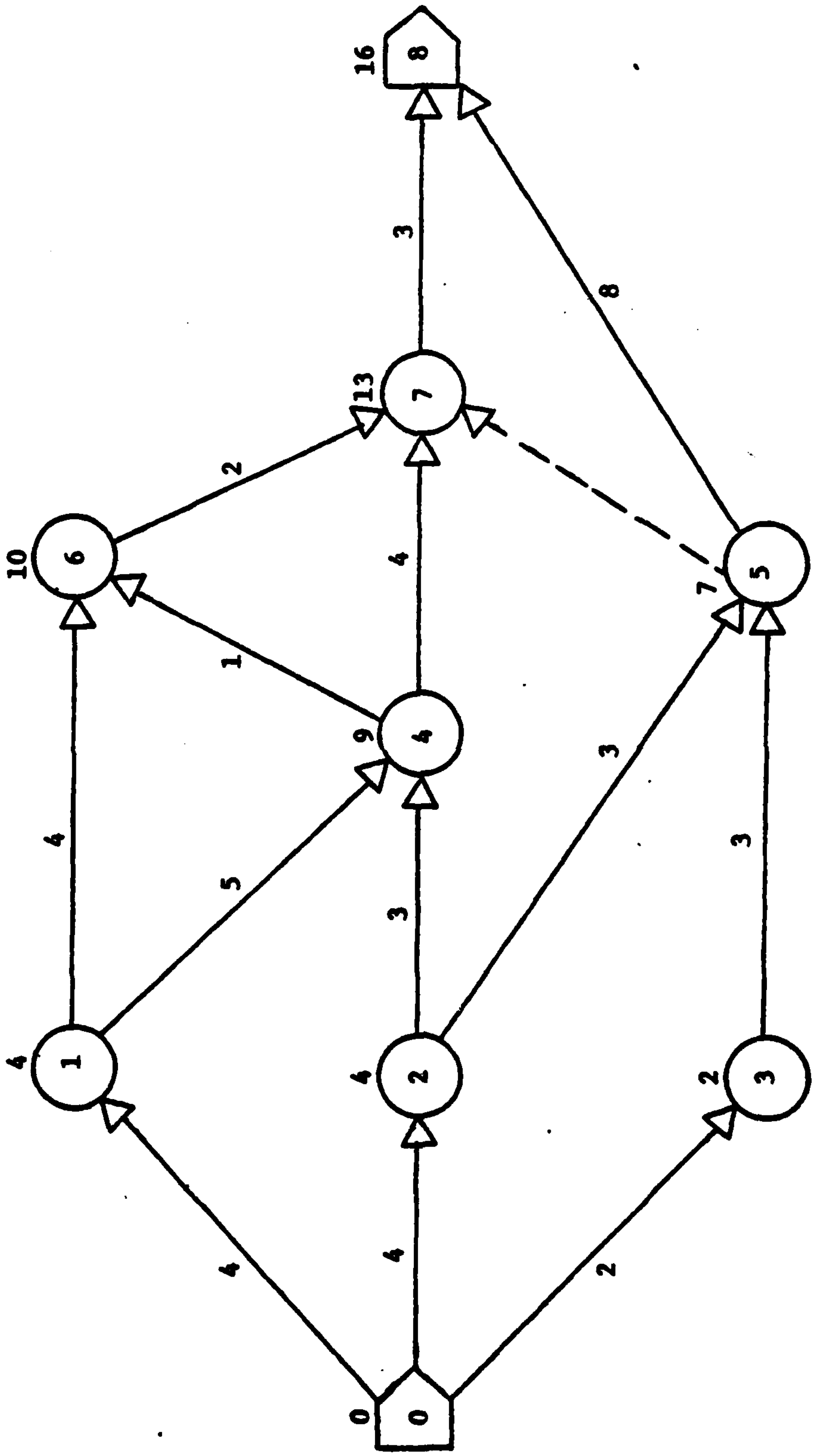


Figure 17-64



As shown in Figure 17-64, the earliest time is 16 units. Having determined the earliest event/project completion time, it is also of interest to find the LATEST TIME for event and project completion. This is the latest time a particular event can occur without delaying completion of the project (in this case, beyond 16 units). The latest completion time for each event is determined by working "backwards" through the network from the sink to the source nodes. The latest time is given by:

$$L_i = \min_j (L_j - t_{ij})$$

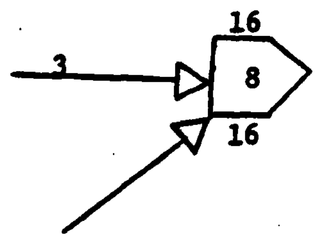
Normally, the latest time is written under the node. For example:

Figure 17-65

<u>Node 8 (Sink)</u>

$L_i = E_i$

$L_8 = E_8 = \underline{16}$



(Sink Node)

As another example:

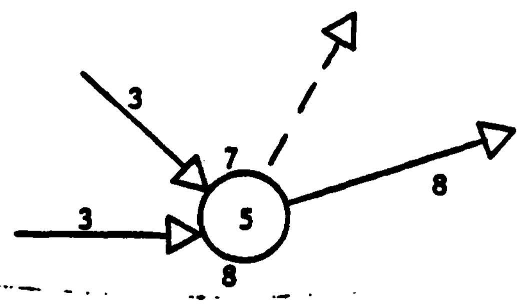
<u>Node 5</u>

$L_5 = L_7 - t_{5,7}$

$= 13 - 0 = 13$

$L_5 = L_8 - t_{5,8}$

$= 16 - 8 = \underline{8}$ (min)



Inspection of the network shows that for certain events (nodes), $E_i = L_i$, the earliest and latest event completion times are the same. This means that there is no "SLACK TIME" available. If the event is not completed on time, the project completion time will also slip. For other events, the latest time exceeds the earliest time. At these events, a slip in the event completion time of $L_i - E_i$ can be incurred without impacting the completion time of the project. This difference, i.e., $L_i - E_i$, is called EVENT SLACK. Events having zero slack are said to be CRITICAL. The path through the network from the source to the sink connecting critical events is called the CRITICAL PATH. It is the <u>longest</u> path through the network (in this case: 0-1, 1-4, and 7-8). The activities defining this path (0-1, 1-4, 4-7, and 7-8) are called CRITICAL ACTIVITIES. This information can be summarized in a table to facilitate project system management and control. An example is presented in Table 17-28. The table is particularly useful in that it highlights the critical activities and events and indicates which activities can be slipped without impact on the project completion time. Figure 17-66 summarizes these computations.

Figure 17-66

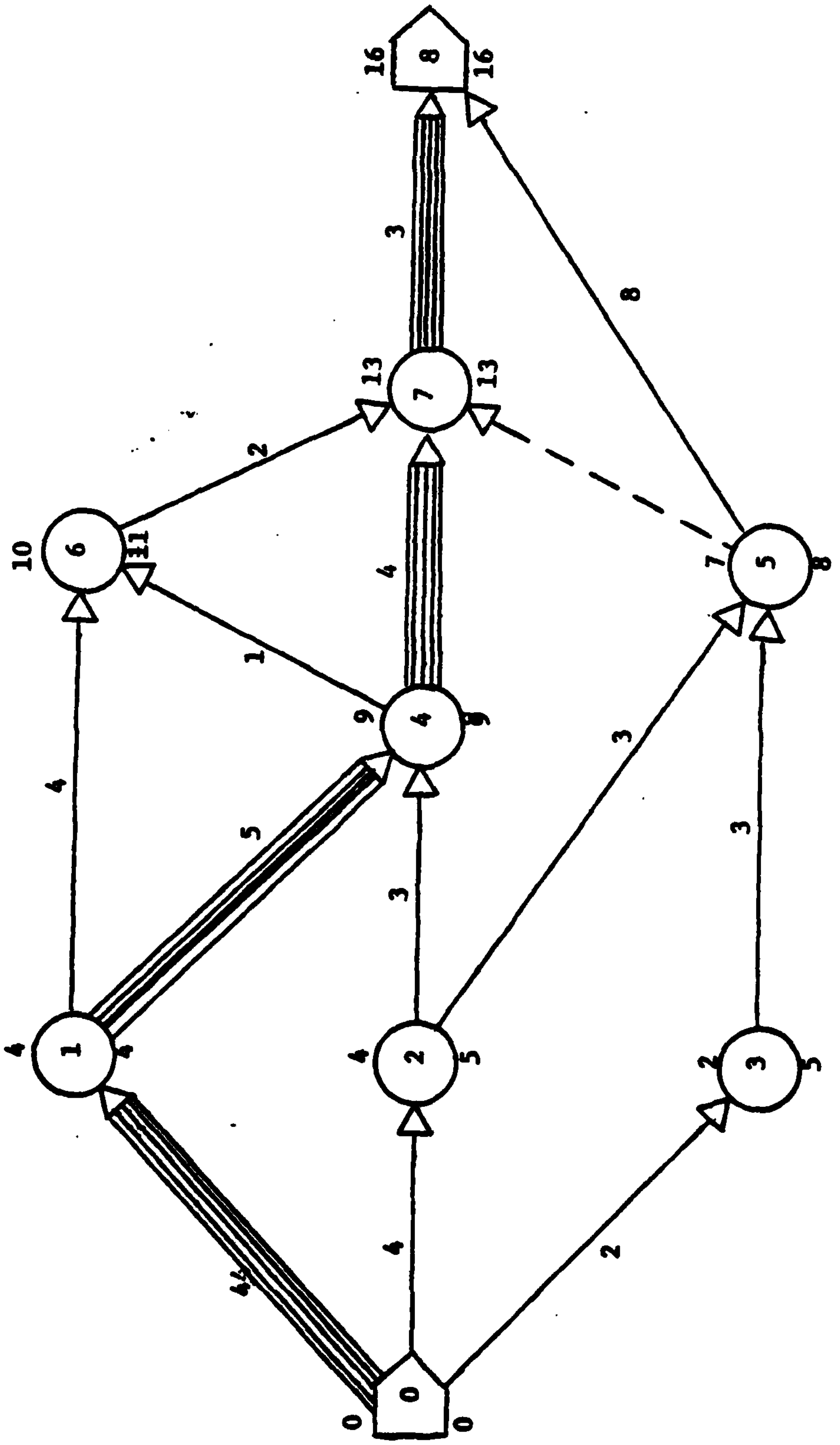


Table 17-28

EVENT	ACT'S	EST. TIME	EARLY		LATE		SLACK	CRIT.
			START	CMPLT.	START	CMPLT.		
0	-	-						
1	0 - 1	4	0	4	0	4	0	X
2	0 - 2	4	0	4	1	5	1	
3	0 - 3	2	0	2	3	5	3	
4	1 - 4	5	4	9	4	9	0	X
	2 - 4	3	4	7	6	9	2	
5	2 - 5	3	4	7	5	8	1	
	3 - 5	3	2	5	5	8	3	
6	1 - 6	4	4	8	7	11	3	
	4 - 6	1	9	10	10	11	1	
7	4 - 7	4	9	13	9	13	0	X
	5 - 7	0	7	7	13	13	5	
	6 - 7	2	10	12	11	13	1	
8	5 - 8	8	7	15	8	16	1	
	7 - 8	3	13	16	13	16	0	X

17.3 PROGRAM EVALUATION AND REVIEW TECHNIQUE (PERT)

PERT was developed by a team of consultants from Lockheed, the Navy, and the firm of Booz, Allen and Hamilton in 1958 (during the Polaris project). It was developed to help project managers deal with an extremely critical time constraint. Prior to that time, completion times on similar projects commonly averaged 30-40% greater than the earliest time estimates, while costs commonly ran two-three times the earliest estimates. Perhaps the most distinctive feature of PERT is its statistical treatment of the estimated activity time, i.e., the activity time estimate is treated as a random variable. It relaxes the assumption of certainty used in CPM and other project management techniques. Activity times are represented by a mean and associated variance. By employing a statistical approach, the project manager can determine the probability of completing a particular activity or event by some scheduled completion date or other date of interest. To estimate the activity mean and variance, PERT uses the following three time estimates for each activity:

a_{ij} = <u>OPTIMISTIC TIME</u>: The minimum reasonable time in which the activity can be completed (there is only a 0-5% chance that the activity can be completed in a shorter period of time)

m_{ij} = <u>MOST LIKELY TIME</u>: The best guess of the time required to complete the activity. This would be the estimate used if the deterministic assumption were made. This is the modal value of the distribution of activity times.

b_{ij} = <u>PESSIMISTIC TIME</u>: The maximum reasonable period of time in which the activity can be completed (there is only a 0-5% chance that the activity will take a longer period of time to complete).

In PERT, the estimated activity completion time is assumed to have a BETA distribution (unless other/historical information suggests otherwise). The BETA distribution was selected because:

1. it is extremely flexible

2. it can take on the variety of forms that typically arise in project activity durations

3. it has finite end points which limit the possible activity times to the area between the a_{ij} and b_{ij} values, and <u>primarily</u>

4. in the simplified version used in PERT, it permits the straightforward computation of the mean, variance, and standard deviation.

The mean is given by: $$t_{e_{ij}} = \frac{a_{ij} + 4m_{ij} + b_{ij}}{6}$$

The variance is given by: $$V_{ij} = \frac{(b_{ij} - a_{ij})^2}{36}$$

To compute the EARLIEST and LATEST event times under PERT, we proceed as before, but use the t_e values instead of the deterministic activity time estimate:

$$E_j = \max_i (E_i + t_{e_{ij}})$$

and

$$L_i = \min_j (L_j - T_{e_{ij}})$$

Again the SLACK for event (i) is given by $L_i - E_i$. It should be noted that these computations assume that the activities in the network are statistically independent. Just as the earliest event time is given by the sum of the t_e values along the longest path from the origin to the event node, the variance in that event time is given by the sum of the associated activity variance values along the <u>same</u> path. Consider the following network:

Figure 17-67

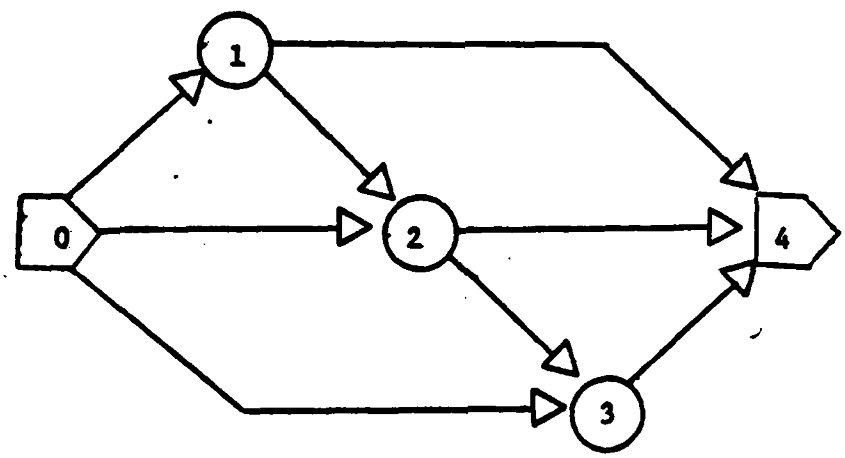


Table 17-29

ACTIVITY	a	m	b	t_e	V
0 - 1	1	5	9		
0 - 2	1	3	11		
0 - 3	1	8	9		
1 - 2	1	1	1		
1 - 4	2	4	6		
2 - 3	1	2	9		
2 - 4	1	4	7		
3 - 4	4	5	12		

We compute the mean (t_e) and variance (V) with the respective formula:

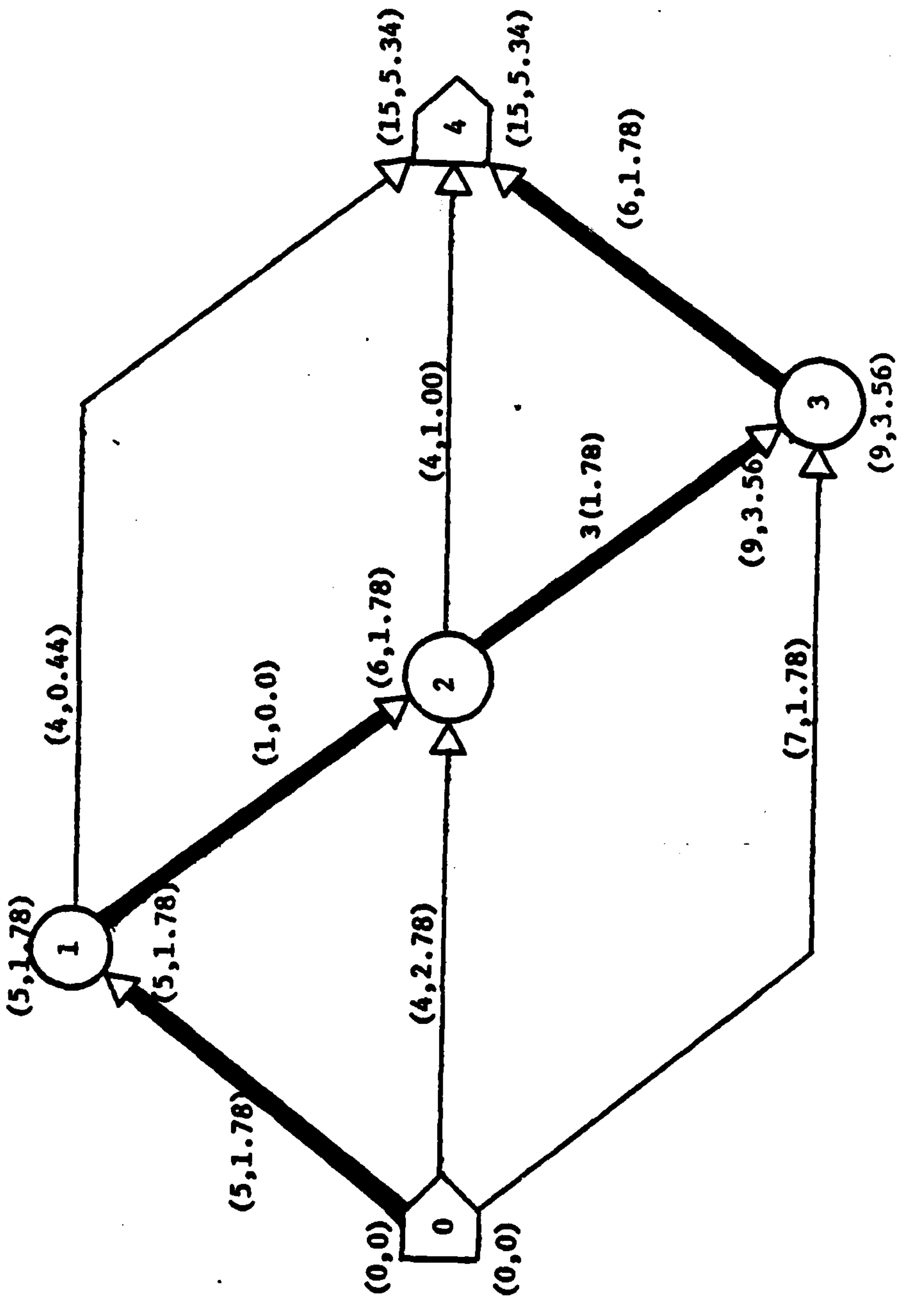
$$T = \frac{a + 4m + b}{6} \quad \text{and} \quad V = \frac{(b - a)}{36}$$

Table 17-30

ACTIVITY	a	m	b	t_e	V
0 - 1	1	5	9	5	1.78
0 - 2	1	3	11	4	2.78
0 - 3	1	8	9	7	1.78
1 - 2	1	1	1	1	0.0
1 - 4	2	4	6	4	0.44
2 - 3	1	2	9	3	1.78
2 - 4	1	4	7	4	1.00
3 - 4	4	5	12	6	1.78

Figure 17-68 shows the network with all acti-ities and events indicating their respective mean values and associated variances as (t_e, V).

Figure 17-68



0
1
2
3
4
(0,0)
(0,0)
(5,1.78)
(5,1.78)
(5,1.78)
(4,2.78)
(7,1.78)
(1,0.0)
(4,0.44)
(6,1.78)
3(1.78)
(9,3.56)
(9,3.56)
(4,1.00)
(6,1.78)
(15,5.34)
(15,5.34)

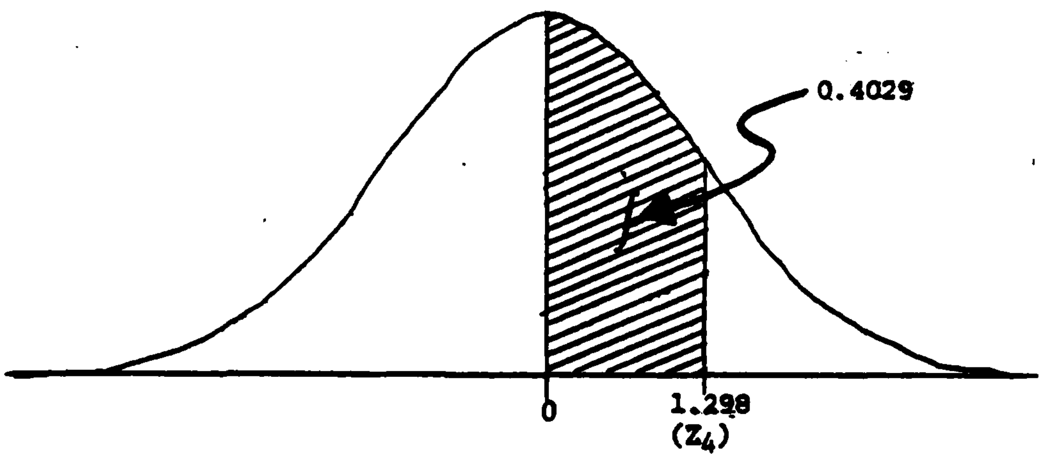
Because PERT treats the estimated activity time as a random variable, it is possible to make statements about the probability of completing an event on or before some scheduled date. By employing the Central Limit Theorem, we can assert that the actual time of <u>event</u> occurrence is NORMALLY distributed with a mean of E_i and a variance of V_i. Therefore, if we define S_i as some scheduled date for event completion, then:

$$Z_i = \frac{S_i - E_i}{\sqrt{V_i}}$$

Suppose, for example, the project manager wants to know the probability of completing the project system illustrated in Figure 17-68 in 18 days or less. We have previously determined that $E_4 = 15$ and $V_4 = 5.34$. Therefore:

$$Z_4 = \frac{18 - 15}{\sqrt{5.34}} = \frac{3}{2.311} = 1.298$$

Figure 17-69



Therefore, for $S_4 = 18$, we can assert that $P(E_4 \le 18) = 0.500 + 0.4029 = 0.9024$, or there is a 90% chance that the project will be completed in 18 days or less.

CPM and Time - Cost Analysis

The original versions of PERT and CPM differ in at least two important ways:

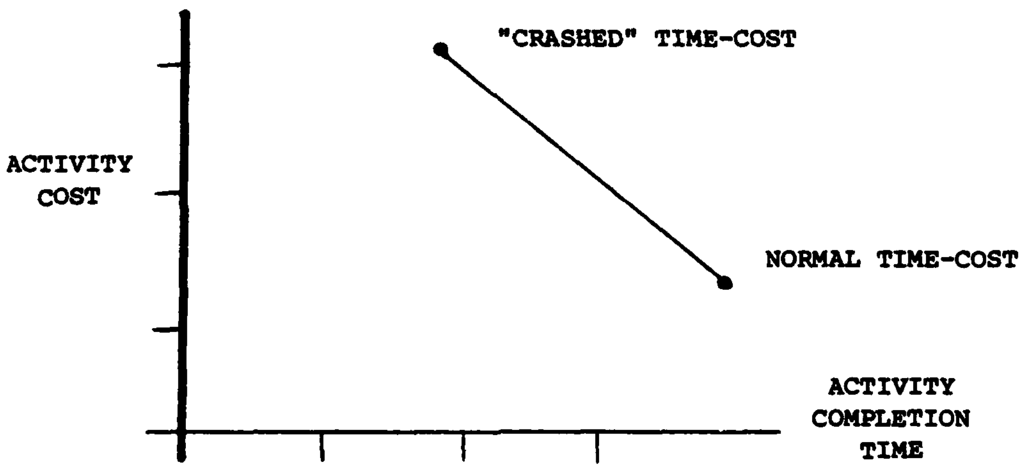
1. First, CPM assumes that activity times are deterministic.

2. Second, rather than focusing primarily or even exclusively on the time factor, CPM places equal emphasis on both time and cost.

The network scheduling logic is the same for both CPM and PERT. The emphasis on EARLY/LATE START and EARLY/LATE FINISH times for individual activities comes from CPM.

CPM assumes that there is a relationship between the time it takes to complete an activity and the cost for that activity, i.e., the completion time can be reduced by some amount by allocating more resources to the activity. In most cases, this relationship is assumed to be linear:

Figure 17-70



	NORMAL	CRASHED
TIME	NT_{ij}	CT_{ij}
COST	NC_{ij}	CC_{ij}

The slope of the time - cost function expresses the MARGINAL EFFECTIVENESS or UTILITY of allocating additional resources to reduce the activity time. Using the notation introduced in the above table, the slope is given by:

$$S_{ij} = \frac{CC_{ij} - NC_{ij}}{NT_{ij} - CT_{ij}}$$

Consider, for example, the system project data summarized in the following table:

Table 17-31

ACTIVITY	NT_{ij}	CT_{ij}	NC_{ij}	CC_{ij}	S_{ij}	CRITICAL
0 - 1	2	1	10	13	3	
0 - 2	1	1	6	6	-	X
0 - 3	4	2	25	33	4	
1 - 3	1	1	7	7	-	
1 - 4	5	2	30	36	2	
2 - 3	2	1	12	19	7	
2 - 5	7	3	35	55	5	X
3 - 4	3	1	15	17	1	
3 - 5	2	1	8	14	6	
3 - 6	3	1	10	16	3	
4 - 6	4	2	14	18	2	
5 - 6	6	2	40	56	4	X

Under normal conditions then, the critical path is 0 - 2 - 5 - 6 with a normal total time of 14 and a normal total cost of 212.

It seems apparent that if the project manager wants to expedite the project completion by allocating additional resources, the first activities to be considered are those currently on the critical path, i.e., those that are currently constraining the project:

ACTIVITY	S	
0 - 2	-	(can't be crashed)
2 - 5	5	
5 - 6	4	

Of these critical activities, we would want to crash activity 5 - 6 first because it has the least expensive unit cost @ 4. But how far can activity 5 - 6 be crashed, i.e., all the way from 6 to 2, or only some portion of the total "crash potential"? We crash the most cost-effective or efficient critical activity until it is no longer critical, or by the entire crash potential amount if the activity remains critical. In this particular case, activity 5 - 6 can be reduced by three time units at an additional cost of 3 @ 4 = 12. At that point, activity 4 - 6 becomes critical and can

Figure 17-71

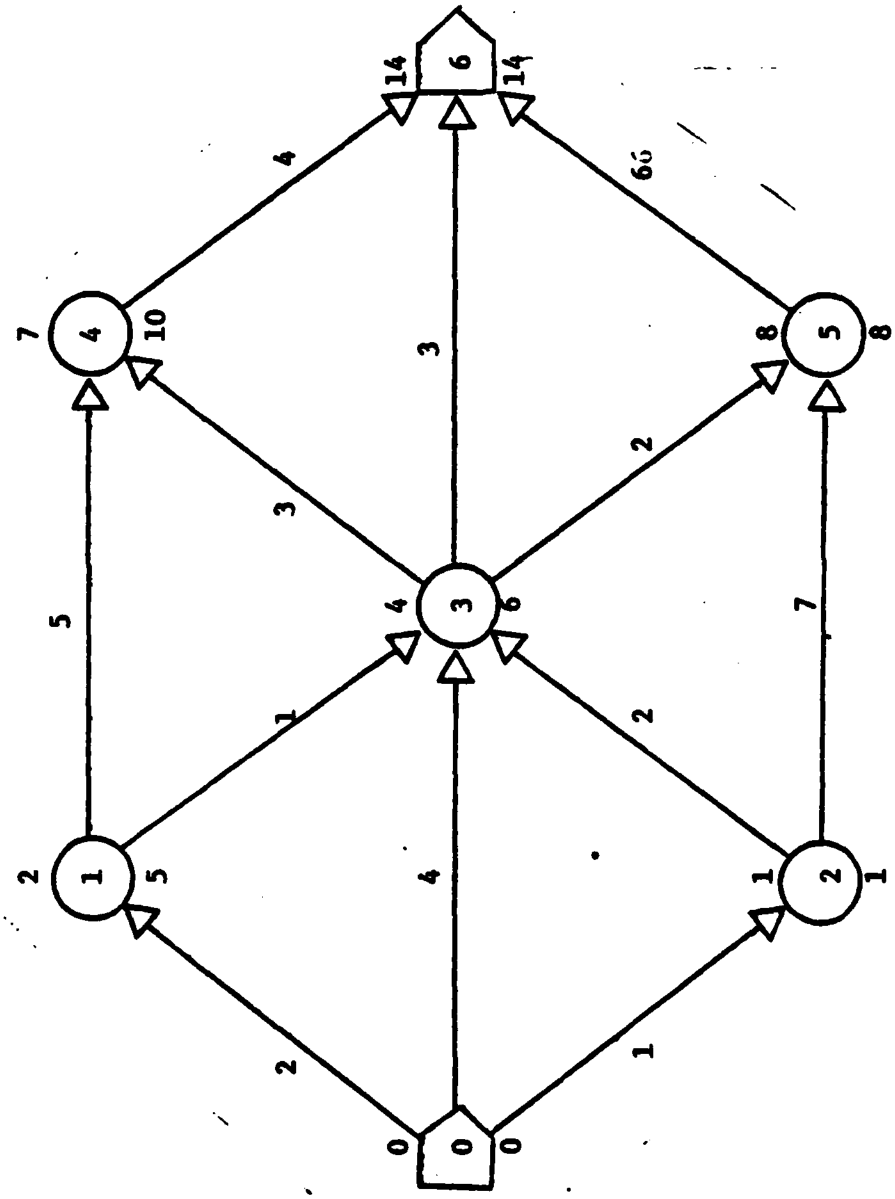
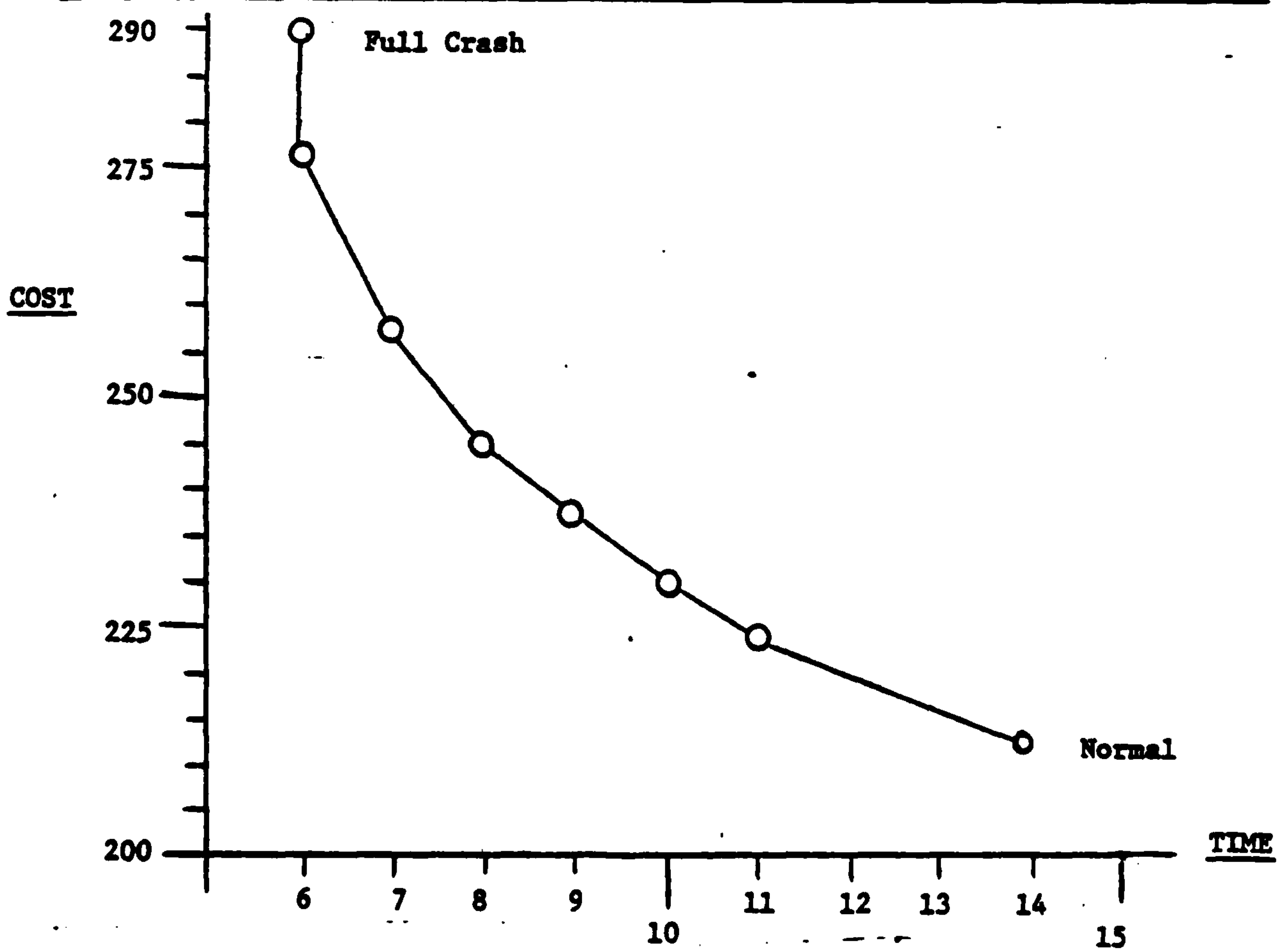


Figure 17-72

ITERATION	ACTIVITY	AMOUNT CRASHED		TOTAL	
		TIME	COST	TIME	COST
Normal				14	212
1	5 - 6	3	12	11	224
2	4 - 6	1	2	10	230
	5 - 6	1	4		
3	4 - 6	1	2	9	237
	2 - 5	1	5		
4	2 - 5	1	5	8	245
	3 - 4	1	1		
	1 - 4	1	2		
5	2 - 5	1	5	7	257
	3 - 4	1	1		
	0 - 3	1	4		
	1 - 4	1	2		
6	1 - 4	1	2	6	277
	0 - 1	1	3		
	0 - 3	1	4		
	3 - 5	1	6		
	2 - 5	1	5		
7	No further reduction is possible				



be crashed at a cost of 2/time unit. To determine the earliest possible completion time under a "full crash" program, we simply continue to crash the most cost-effective critical activity until it is no longer critical. For even moderate-sized project networks, it is difficult to systematically crash a project system without the assistance of a computer code. Figure 17-72 summarizes a full crash program for this project. Note that it is not necessary to crash every activity by the full amount in order to achieve the maximum possible reduction in completion time. In this case, crashing the network from a total cost of 277 to 290 doesn't result in any further reduction in project completion time.

REFERENCES

1. Anderson, D. R., D. J. Sweeney, and T. A. Williams. An Introduction to Management Science: Quantitative Approaches to Decision Making. St Paul, MN: West Publishing Co., 1976.

2. Budnick, F. S., R. Mojena, and T. E. Vollmann. Principles of Operations Research for Management. Homewood, IL: Richard D. Irwin, Inc., 1977.

3. Chase, R. B., and N. J. Acquilano. Production Operations Management: A Life Cycle Approach. Homewood, IL: Richard D. Irwin, Inc., 1973.

4. Cleland, D. I., and W. R. King. Systems Analysis and Project Management. New York, NY: McGraw-Hill Book Company, 1975.

5. Daellenbach, H. G., and J. A. George. Introduction to Operations Research Techniques. Boston, MA: Allyn and Bacon, Inc., 1978.

6. Heinze, D. Management Science: Introductory Concepts and Applications. Cincinnati, OH: South-Western Publishing Co., 1978.

7. Hillier, F. S., and G. J. Lieberman. Introduction to Operations Research. 2d ed. San Francisco, CA: Holden-Day, Inc., 1974.

8. Levin, R. I., and C. A. Kirkpatrick. Quantitative Approaches to Management. 4th ed. New York, NY: McGraw-Hill Book Company, 1978.

9. Star, M. K. Operations Management. Englewood Cliffs, NY: Prentice-Hall, Inc., 1978.

10. Thompson, G. E., Management Science: An Introduction to Modern Quantitative Analysis and Decision Making. New York, NY: McGraw-Hill Book Company, 1976.

11. Turban, E., and J. R. Meredith. Fundamentals of Management Science. Dallas, TX: Business Publications, Inc., 1977.

12. Wagner, H. M. Principles of Management Science: With Applications to Executive Decisions. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1970.

CHAPTER 18

INVENTORY MODELS

INTRODUCTION

This chapter presents the basic models developed for the management of inventory where inventory will be referred to as the stock of any independent demand item used in the conduct of an organization's mission or operation. Inventory serves many purposes, chief of which include: decoupling successive stages in production, distribution and consumption processes; protection against variations in item demand and/or delivery schedules; and reduction of aggregate inventory costs through establishment of economic purchase order replenishment levels.

Inventory models are used to assist inventory managers in determining when items should be ordered and how large the order should be so as to minimize relevant variable costs, while ensuring that adequate stocks are on hand to meet mission requirements.

EOQ MODEL

The classic economic order quantity or EOQ model, which serves as the foundation for the more complex but realistic inventory models, is used under the following ideal conditions[1]:

1. Demand per time period is known with certainty and is at a constant (linear) rate.

2. Lead time is known with certainty and equals zero.

3. Stockouts are not permissible.

4. Replenishment is instantaneous.

5. Order quantities are always the same size and are not restricted to being discrete.

6. Unit cost for items is constant with no discounts.

7. Single-item, single echelon inventories are assumed.

8. An infinite planning horizon is assumed.

9. Demand, lead time and costs are fixed over time.

[1]These conditions, or assumptions, and the notation used in this chapter are taken from Frank S. Budnick, Richard Mojena, and Thomas E. Vollmann, Principles of Operations Research for Management (Homewood, Il: Richard D. Irwin, Inc., 1977), Chapter 11.

The objective of the EOQ model is to determine the optimum economic order quantity (Q^*) and number of orders (N^*) to place per given time period so as to minimize total inventory costs given by:

$$\text{Total Cost} = \begin{array}{c}\text{Ordering Cost}\\ \text{per Time Period}\end{array} + \begin{array}{c}\text{Carrying Cost}\\ \text{per Time Period}\end{array} + \text{Purchase Cost}$$

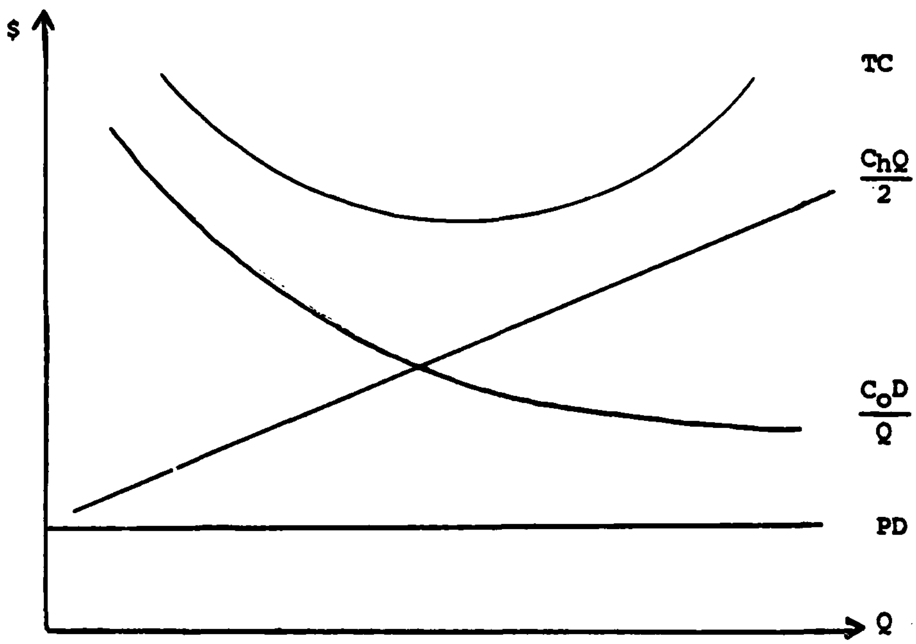
or

$$TC = \frac{C_o D}{Q} + \frac{C_h Q}{2} + PD \qquad (1)$$

where TC = Total inventory cost
D = Demand per time period
C_o = Order cost per order
Q = Order quantity
C_h = Carry or holding cost per unit per time period
P = Purchase price per unit

The behavior of the inventory cost components is illustrated in figure 1.

Figure 1. Inventory Costs



To minimize the total cost equation, we take the first derivative of the expression with respect to Q, set the result equal to zero and solve for Q:

$$\frac{dTC}{dQ} = \frac{dC_oDQ^{-1}}{dQ} + \frac{d\frac{1}{2}C_hQ}{dQ} + \frac{dPD}{dQ}$$

$$\frac{dTC}{dQ} = -\frac{C_oQ}{Q^2} + \frac{C_h}{2} = 0$$

$$-C_oD + \frac{Q^2C_h}{2} = 0$$

$$Q^2 = \frac{2C_oD}{C_h}$$

$$Q^* = \sqrt{\frac{2C_oD}{C_h}} \qquad (2)$$

Note that the purchase cost component falls out of the final result since the price per unit was assumed fixed with respect to the order quantity. Other important results for this model are:

$$N^* = \frac{D}{Q} \qquad (3)$$

$$t^*_c = \frac{Q^*}{D} \qquad (4)$$

where N^* = the optimum number of orders to place per time period

t^*_c = the inventory cycle time or time between order replenishment and depletion.

EXAMPLE 18-1.

Under the consolidated administrative airlift program, 109 T39 aircraft, operated by MAC, are located throughout the CONUS to provide pilot training and passenger movement capability. Main gear tires for these aircraft are procured and managed at the Ogden Air Logistics Center. The item manager is interested in establishing an inventory policy for these tires (that is, how many tires to order each time an order is placed and how many orders should be placed each year). The annual requirements are estimated at 2250 tires. It costs $25 to place an order and the cost of carrying one tire in inventory for one year is $5. Lead time is zero days since the supplier can deliver tires to the Ogden ALC on the order date. The purchase cost is $50 per tire.

For this problem, Q^* is determined from equation (2):

$$Q^* = \sqrt{\frac{2(25)(2250)}{5}} = 150 \text{ tires}$$

The optimum number of orders to place is given by equation (3):

$$N^* = \frac{2250}{150} = 15 \text{ orders per year}$$

The cycle time is provided from equation (4):

$$T_c = \frac{150}{2250} = .067 \text{ year or } 24.333 \text{ days}$$

The total inventory cost of this policy is given by equation (1):

$$TC = \frac{25(2250)}{150} + \frac{5(150)}{2} + 50(2250) = \$113,250$$

EOQ Model with Deterministic Lead Time

In most inventory systems, the time between order placement and receipt is not zero days, as assumed in the classic EOQ model. As such, order placement must be timed so that the order arrives when the inventory on hand is depleted. The point in time at which the order is made is referred to as the reorder point which is commonly expressed as a predetermined inventory level. With deterministic lead times, the reorder point equation is:

$$R^* = D \cdot t_L - \left[\frac{t_L}{t_c}\right] Q^* \qquad (5)$$

where t_L = lead time
D = demand (in lead time dimensions)
$D \cdot t_L$ = demand during lead time
$\left[\frac{t_L}{t_c}\right]$ = integer part of t_L/t_c or the number of complete cycles during a lead time period.

EXAMPLE 18-2

For the aircraft tire problem, suppose that the lead time is 30 days, that is, it takes 30 days from the time the order is placed until the order arrives. From equation (5) we have:

$$R^* = \frac{2250 \text{ tires}}{365 \text{ days}} \cdot 30 \text{ days} - \left[\frac{30 \text{ days}}{24.333 \text{ days}}\right] 150 \text{ tires}$$

$$= 184.9 \text{ tires} - (1)(150 \text{ tires})$$

$$= 34.9 \text{ tires}$$

The optimal policy is, therefore, to place an order for 150 tires (Q*) when the inventory on hand drops to 35 tires.

EOQ Model with Backorders

The classic EOQ model can be extended to explicitly consider the case where shortages are permissible. With shortages or backorders allowed, all demands will continue to be satisfied but not in the inventory cycle in which they were made. Although difficult to measure, a backorder cost should be included in the total inventory cost equation and policies developed so as to balance this cost against the order and carry cost components. The total cost equation can be expressed as:

$$TC = \frac{C_o D}{Q} + \frac{(Q-S)^2 C_h}{2Q} + \frac{S^2 C_s}{2Q} + PD \qquad (6)$$

where S = units short that must be backordered during each inventory cycle.
C_s = shortage (backorder) cost per unit per time period.

The policy variables, Q* and S*, are determined by differentiating the total cost equation with respect to Q and S, setting the resulting partial derivatives equal to zero and solving for Q* and S* simultaneously. The results are:

$$Q^* = \sqrt{\frac{2C_o D}{C_h} \left(\frac{C_h + C_s}{C_s}\right)} \qquad (7)$$

$$S^* = \sqrt{\frac{2C_o D C_h}{C_s C_h + C_s^2}} \qquad (8)$$

Other results for the EOQ model with back orders are:

$$L^* = Q^* - S^* \qquad (9)$$

$$t_1 = \frac{L^*}{D} \qquad (10)$$

$$t_2 = \frac{S^*}{D} \tag{11}$$

$$t_c = t_1 + t_2 = \frac{Q^*}{D} \tag{12}$$

$$R^* = Dt_L - \left[\frac{t_L}{t_c}\right] Q^* - S^* \tag{13}$$

where L^* = real (maximum) inventory level
t_1 = time during an inventory cycle when inventory is held.
t_2 = time during an inventory cycle when inventory is backordered.

EXAMPLE 18-3.

For the aircraft tire problem, suppose that backorder costs are \$40 per tire per year. The policy variables and other results are as follows:

$$Q^* = \sqrt{\frac{2(25)(2250)}{5}\left(\frac{5+40}{40}\right)} = 159.1 \text{ tires in each order}$$

$$S^* = \sqrt{\frac{2(25)(2250)(5)}{(5)(40)+(40)^2}} = 17.7 \text{ tires backordered each cycle}$$

$$N^* = \frac{2250}{159.1} = 14.1 \text{ orders per year}$$

$$t_c = \frac{159.1}{2250} = .0707 \text{ year or } 25.8 \text{ days}$$

$$L^* = 159.1 - 17.7 = 141.4 \text{ tires}$$

$$t_1 = \frac{141.4}{2250} = .0628 \text{ year or } 22.9 \text{ days}$$

$$t_2 = \frac{17.7}{2250} = .0079 \text{ year or } 2.9 \text{ days}$$

$$R^* = \frac{2250 \text{ tires}}{365 \text{ days}} \cdot 30 \text{ days} - \left[\frac{30 \text{ days}}{25.8 \text{ days}}\right] 159.1 \text{ tires} - 17.7 \text{ tires}$$

$$= 184.9 - 159.1 - 17.7 = 8 \text{ tires}$$

$$TC = \frac{25(2250)}{159.1} + \frac{5(159.1 - 17.7)^2}{2\ (159.1)} + \frac{(17.7)^2(40)}{2\ (159.1)}$$

$$+ 50(2250)$$

$$= \$113,207.11$$

The results suggest that approximately 14 orders should be placed each year, each consisting of 159 tires. The number of backorders planned for each cycle are 18 tires and an order should be placed when the inventory drops to 8 tires. Note that orders placed in cycle k arrive in cycle k+2. The total number of backorders per year are:

$$S^*N^* = 17.7\ (14.1) = 249.57 \text{ units}$$

and the total days that backorders exist during the year are:

$$t_2N^* = 2.9\ (14.1) = 40.9 \text{ days}$$

EOQ Model with Quantity Discounts

For certain items, suppliers offer price discounts if buyers make purchase orders of large quantities. With these discounts, per unit purchasing cost and total ordering costs decrease while carrying costs will increase. The following example will illustrate a simple total cost comparison procedure for determining if quantity discounts should be taken.

EXAMPLE 18-4

Suppose for the aircraft tire problem, the supplier offers the following discount schedule:

Order Quantity	Discount	Purchase Price per Unit
$0 < Q < 200$	None	\$50.00
$200 \leq Q < 800$	2%	49.00
$800 \leq Q$	3%	48.50

The first step is to compute Q^* from equation (2) and the associated total inventory cost for this policy using the appropriate unit price from the table:

$$Q^* = \sqrt{\frac{2(25)(2250)}{5}} = 150 \text{ tires}$$

Since Q^* is less than 200 units, the purchase price is \$50.00 per tire and the total inventory cost is given by equation (1):

$$TC = \frac{25(2250)}{150} + \frac{5(150)}{2} + 50(2250) = \$113,250$$

The next step is to recompute the total inventory cost using the next lower discount price of $49.00. Note that the only way total costs can decrease is if the buyer purchases the minimum quantity Q associated with the 2% discount because of the shape of the total cost curve. With Q = 200 and P = $49.00, the total inventory cost from equation (1) is:

$$TC = \frac{25(2250)}{200} + \frac{5(200)}{2} + 49(2250) = \$111,031.25$$

The above step is repeated through the entire discount schedule. For the 3% discount, Q is set equal to 800 and P = $48.50. The total cost for this policy is:

$$TC = \frac{25(2250)}{800} + \frac{5(800)}{2} + 48.5(2250) = \$111,195.31$$

The last step is to select that policy which corresponds to the minimum total cost. For this example, the minimum total cost is $111,031.25 which is associated with an order quantity of 200 tires and a purchase price of $49.00.

EOQ Model with Stochastic Demand

In most inventory systems, including those used in the USAF, the demand and lead time parameters are stochastic rather than deterministic, as assumed by the classic EOQ model. For the fixed order quantity (EOQ) model without backorders, demand variability during lead time and/or lead time variability could result in an unprogrammed shortage occurring prior to inventory replenishment. To protect against these shortages, a safety level quantity or safety stock is added to the demand during lead time quantity for computation of the reorder point. The safety stock is commonly established as a policy variable derived from management assessment of the risk of incurring backorders and knowledge of the distribution of demand during lead time parameter.

This section will discuss the stochastic demand during deterministic lead time situation. It is recognized that lead time may also be stochastic although there are some management controls over this parameter to reduce the uncertainty. If both demand during lead time and lead time are stochastic, the distribution of demand during lead time is difficult to mathematically determine since a convolution of the two parameters is required to generate the function. However, estimates of this distribution can be made by experimentally observing the interaction of demand and lead time through a simulation model of the two processes.

The EOQ model and reorder point equations are modified as follows to incorporate stochastic demand during deterministic lead time:

$$Q^* = \sqrt{\frac{2C_o\bar{d}}{C_h}} \tag{14}$$

where $\bar{d}$ = average daily demand determined from the demand distribution

C_h = carry cost per unit per day

Once Q^* is computed, the reorder point is computed independently as follows (this equation assumes that demand during lead time can be described by a normal distribution).

$$R^* = \bar{d}_L + Z_{1-\alpha}\,\sigma_{\bar{d}_L} \tag{15}$$

where $\bar{d}_L$ = average demand during lead time

$\sigma_{\bar{d}_L}$ = the standard deviation of demand during lead time

$Z_{1-\alpha}$ = number of standard normal deviates needed to achieve a cumulative probability (service level) of $1-\alpha$

$Z_{1-\alpha}\sigma_{\bar{d}_L}$ = safety stock or safety level quantity

EXAMPLE 18-5.

For the aircraft tire problem suppose that the average daily demand is now 6 tires per day, the order cost is \$25 per order and the carry cost is \$5 per unit per year. The economic order quantity is computed from equation (14) (note that the per unit order cost is divided by 365 to convert all dimensions to days):

$$Q^* = \sqrt{\frac{2(25)6}{\left(\frac{5}{365}\right)}} = 147.99 \text{ or } 148 \text{ tires}$$

This result suggests that each order will be for 148 tires and

$$N^* = \frac{\bar{d} \times 365}{148} = 14.8 \text{ or } 15 \text{ orders will be placed per year.}$$

For computation of the reorder point, assume that management establishes a stock out risk of 10%. That is, if 100 random observations were taken during the resupply time period, in 10 of those observations at least one shortage would be expected. The complement of the stock out risk is 90% (1-10%) which is referred to as the service level, that is, in 90 out of 100 random observations taken during the resupply time period, no backorders or stock outs would be expected to incur. Thus $\alpha = .10$ and $1-\alpha = .90$.

The following information has been collected for computation of $\bar{d}_L$ and $\sigma_{\bar{d}_L}$:

Class Intervals for Demand During Lead Time	Class Mark: Demand During Lead Time (d_{L_i})	Frequency of Occurrence $F(d_{L_i})$
7 but under 9	8	.10
9 but under 11	10	.20
11 but under 13	12	.40
13 but under 15	14	.20
15 but under 17	16	.10

The average demand during lead time (mean of a frequency distribution) is computed from the following statistical formula:

$$\bar{d}_L = \sum_i d_{L_i} f(d_{L_i})$$

$$= 8(.10) + 10(.20) + 12(.40) + 14(.20) + 16(.10) = 12$$

The standard deviation of demand during lead time (standard deviation of a frequency distribution) is computed from the following statistical formula:

$$\sigma_{\bar{d}_L} = \left[\sum_i (\bar{d}_{L_i} - \bar{d}_L)^2 f(d_{L_i})\right]^{1/2} =$$

$$\left[(8-12)^2(.10)+(10-12)^2(.20)+(12-12)^2(.40)+(14-12)^2(.20)+(16-12)^2(.10)\right]^{1/2} = 2.19$$

Given $\alpha = .10$, then $Z_{1-\alpha} = Z_{.90} = 1.282$ and the reorder point is determined from equation (15):

$$R^* = 12 + 1.282\ (2.19) = 14.8 \text{ or approximately 15 tires.}$$

Therefore, an order for 148 tires will be placed each time the inventory falls to 15 units on hand. Note that the safety stock is 2.8 or approximately 3 tires which is expected to provide a service level of 90% or alternatively, with a safety level of 3 tires, the probability of having less than zero tires on hand when an order is received is less than 10%.

Periodic Review Inventory Systems

The models discussed up to this point are used in continuous review inventory systems where on hand inventory is continuously reviewed and an order is placed for a fixed quantity Q^* when the on hand level is observed to reach the reorder point R^*. Periodic review inventory systems require the review of on hand inventory levels every t_c time units at which time an order is placed for a variable quantity, q, to bring the inventory level up to a predetermined replinishment level, M.

For the periodic review inventory system with stochastic demand during deterministic lead time, Q is computed from equation (14) and t_c is computed as follows:

$$t_c = \frac{Q^*}{\bar{d}} \qquad (16)$$

From the example problem presented in the last section, this equation gives:

$$t_c = \frac{148}{6} = 24.67 \text{ days}$$

The periodic review system requires more safety stock than the continuous review system. Safety stock for the continuous review system was computed using the probability distribution of demand during lead time. Safety stock for the periodic review system, however, is computed using the probability distribution of demand over a longer time period, namely, lead time plus the review period. The replenishment level, which accounts for this increased safety stock, is:[2]

$$M = \bar{d}_{t_L+t_c} + Z_{1-\alpha}\, \sigma_{\bar{d}_{t_L+t_c}} \qquad (17)$$

where $\bar{d}_{t_L+t_c} = \bar{d}(t_L + t_c)$

$Z_{1-\alpha}\sigma_{\bar{d}_{t_L+t_c}}$ = safety stock

$\sigma_{\bar{d}_{t_L+t_c}} = \sigma_{\bar{d}}(t_L+t_c)^{\frac{1}{2}}$

To illustrate the application of equation (17), let the average daily demand be 6 units per day, the lead time be 2 days and the standard deviation of average daily demand be 1 unit per day for the example problem. If $\alpha = .10$, $Z_{1-\alpha} = Z_{.90} = 1.282$ and the replenishment level is:

$$\begin{aligned} M &= 6(24.67 + 2) + 1.282(1)(24.67 + 2)^{\frac{1}{2}} \\ &= 160 + 6.6 \\ &= + 166.6 \text{ or approximately 167 tires} \end{aligned}$$

The order quantity for this sytem cannot be a fixed quantity as it was for the continuous review system. The reorder policy must be adaptive and based upon the on-hand inventory (I) at the time of the review. The reorder quantity is given by: q = M-I.

[2]Richard B. Chase and Nicholas J. Aquilano, Production and Operation Management (Homewood IL: Richard D. Irwin, Inc., 1977) pp. 388-389.

The Air Force EOQ Model

The EOQ model used in the Air Force supply system is applied to items which cannot be economically repaired by a field or depot maintenance activity. As covered in Chapter 11, AFM 67-1, these items include consumables, minor parts, components, tools and hardware. The model is derived from a total cost equation given as:

$$TC = \frac{C_O DDR}{Q} + \frac{I \cdot P \cdot Q}{2} \tag{18}$$

where DDR = daily demand rate = $\frac{\text{cummulative recurring demands}}{\text{current date - date of first demand}}$

I = 50% carry cost rate

C_O = \$5 per order

The resulting EOQ model is

$$EOQ = \frac{4.4\sqrt{DDR \times VSO \times P}}{P} \tag{19}$$

where VSO is a variable stockage objective days weighting factor. The safety level quantity is given by:

$$SLQ = C\sqrt{3(O \;\&\; STQ)} \tag{20}$$

where C = the number of standard deviations to be covered by the SLQ which is normally set equal to 1 giving a 84% service level.

O & STQ = the order and ship time quantity

Finally, the reorder point is given by: ROP = SLQ + O & STQ (21)

Recoverable Item Model

The continuous review inventory system model can be easily adapted to describe the recoverable item control system used by the Air Force. In fact, the modification is only a redefinition of lead time. Recoverable items are repaired, not disposed of, when they fail. Lead time for such a system is, then, the resupply time. Resupply time is defined to be the time from failure of an item until the item is either repaired at the local level and returned to supply or is replaced by a one-for-one trade transaction with the depot repair activity. The average resupply time, T, is given in abbreviated form by:

$$T = r\,(R_B) + (1-r)\,R_O \tag{22}$$

where R_B = average repair time at the base

R_O = average replacement time from the depot (order and ship time)

r = probability that the item will be repaired at base level

$1-r$ = Not Reparable This Station (NRTS) rate.

The repair cycle quantity (RCQ) is then computed as follows:

$$RCQ = DDR \cdot PBR \cdot T \tag{23}$$

where PBR is the percentage of time (decimal equivalent) that a given item is repaired at base level. Finally, the order and ship time quantity is computed as follows:

$$O\ \&\ STQ = DDR\,(1-PBR) \cdot O\ \&\ ST \tag{24}$$

Concluding Remarks

Inventory models which support continuous review and periodic review systems have been presented in this chapter. Most organizations with a substantial investment in inventory find it economical to stratify the items they stock according to some simple classification scheme, such as the "A-B-C" scheme. Items which fall into category A typically make up 10% of the total inventory volume, but account for as much as 70% of total investment. These are high value items and usually require a continuous review system model to control their activity. The remaining items can be controlled adequately with a periodic review system.

The models presented do have shortcomings, chief of which is their limitation to single item inventories. When applied to large multi-item systems, they do not address the interactive nature of such systems. The mathematical models are useful for gaining some valuable theoretic insights into system behavior and performance, but lack the ability to deal with the enormous complexity found in real systems. To deal with the complex multiple item, multiple echelon systems, we can return to the empirical scientific technique of experimenting with an adequate model of the system. Such a model for experimentation can be developed using the techniques of computer simulation.

REFERENCES

Budnick, Frank S., Mojena, Richard and Vollmann, Thomas E. Principles of Operations Research for Management, Homewood IL: Richard D. Irwin, Inc., 1977.

Chase, Richard B., and Aquilano, Nicholas J. Production and Operations Management, Homewood IL: Richard D. Irwin, Inc., 1977.

__________. USAF Supply Manual: USAF Standard Base Supply System. AFM 67-1, Volume II, Part Two, Chapter 11. Washington: Government Printing Office, 1 January 1979.

CHAPTER 19

QUEUING MODELS

INTRODUCTION

Waiting lines or queues arise in numerous situations within the logistics environment. Essentially, a queue forms whenever people or objects, arriving at a service facility, must wait because the service operation is busy. In these situations where demand for service exceeds the capacity of the server or service facility, the organization or mission incurs direct or at least indirect costs since the time people or objects spend in waiting is unproductive. However, minimizing the cost of waiting is not necessarily an easy matter since increasing the capacity of the service facility to respond to demand can lead to increases in the marginal cost of service. Accordingly, a primary objective in the analysis of queuing situations is to achieve a balance in the tradeoff between the cost of waiting and the cost of service. Many analytical models have been developed to describe certain operating characteristics of the queuing phenomena and are therefore useful to decision makers in the analysis of cost tradeoffs and other aspects of the queuing system. This chapter will survey several of the fundamental analytical techniques.

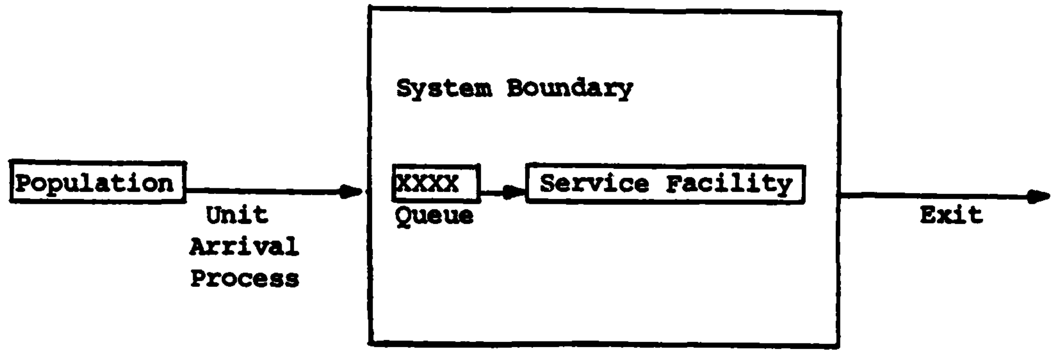
The examples used to illustrate the quantitative models may appear less complex than many of the queuing situations encountered in the logistics environment since the assumptions about the queuing process being described by the models are quite restrictive. In reality, queuing situations can typically be characterized as dynamic, interactive systems with complex probabilistic arrival and servicing patterns. For these intricate waiting line situations, computer simulation, which involves development of a simulation model to artificially reproduce and describe the queuing process, is a viable recourse. However, this does not suggest that the quantitative models have no utility for the decision maker. Their importance is recognized in that model output provides useful information which, in many instances, adequately approximates the operating characteristics of the queuing process being described and which may be useful in the verification of simulation model relationships.

The word "described" was used above, with consideration. All queuing technologies, whether applied mathematics or simulation, are descriptive devices. Their output is not an optimal solution but rather a prediction of the system's plausible future behavior. This prediction can only assess the likelihood that the system will be in any of its possible states at any particular point in time; that is, no model has the capability to predict any future state with certainty.

QUEUING SYSTEMS

The typical structure of a queuing system is given in Figure 19-1.

Figure 19-1. Queuing System



The major components and processes of the queuing system are described as follows:

a. Population--the source of the units arriving at the service facility. This calling source can consist of single or multiple populations having finite or infinite number of units.

b. Unit Arrival Process--the manner by which units arrive at the queuing system.

1. Units can arrive singularly or in bulk and under total, partial or no control.

2. The arrival process can be deterministic or stochastic with dependent or independent arrivals.

c. Queue--units awaiting service.

1. The queue is described by its length, which may be restricted; by the number of lines; and by the relationship of the lines to the service facility.

2. Queue discipline refers to the behavior of arriving units. Arriving units may view the line and decide not to join (balk); join the line and then leave the system without being served (renege); exchange lines in multiple queue systems (jockey); collude with other units in line; and use a variety of decision rules, such as random selection, in selecting a line in multiple queue systems. Arriving units which randomly select a line in multiple queue systems and which do not balk, renege, collude or jockey, are referred to as patient units.

d. Service Facility--that entity in which the process of service is embodied.

1. The service facility may have none (unit self service process), one or multiple servers in parallel, series or both configurations.

2. Multiple servers may be cooperative or uncooperative, meaning that idle servers either assist or do not assist busy servers in the system.

3. The time required to service one unit in the facility may be deterministic or stochastic.

4. Service discipline describes the manner in which units are selected for processing and includes, for example, first-come first-served (FCFS), last-come first-served (LCFS), service in random order (SIRO) and preemptive and nonpreemptive priority disciplines.

Terminology and Notation

a. State of the system--$N(t)$: typically, the number of units in the system at time t.

b. State probabilities--$P_n(t)$: the probability that the system will be in state n at time t; that is, n units in the system at time t.

c. Number of servers--c.

d. Mean arrival rate of units when n units are already in the system--λ_n.

e. Mean service rate for the total system when n units are in the system--μ_n.

Steady State Notation

Because the behavior of a stochastic, dynamic system is difficult to capture in a model consisting only of mathematical and statistical variables and relationships, such efforts always entail the assumption that the stochastic processes generating the probabilistic phenomena can and have attained a condition of steady state. This is a condition of probabilistic equilibrium and indicates only that the probabilities of transitioning from one state to another are independent of initial conditions. Special notation has been developed for steady state conditions. Some of the most frequently used are:

a. State probabilities--P_n.

b. Expected number of units in the system--L_s.

c. Expected number of units in the queue--L_q.

d. Expected number of units in the queue for a busy system--L_b.

e. Expected time in the system--W_s.

f. Expected time in the queue--W_q.

g. Expected time in the queue for a busy system--W_b.

Relationship between L_s, L_q, W_s and W_q

If λ_n is a constant λ for all n, and μ_n is a constant μ for all n, and the symbol "E" represents the expected value or mean, then:

E [number in system] = E [number in queue] + E [number in service]

or

$$L_s = L_q + L_{se} \quad (1)$$

E [time in system] = E [time in queue] + E [time in service]

or

$$W_s = W_q + \frac{1}{u} \quad (2)$$

$$L_s = \lambda W_s \quad (3)$$

$$L_q = \lambda W_q \quad (4)$$

Once any three of the seven variables have been measured or estimated, these equations may be used to provide the solution for the remaining four variables.

FUNDAMENTAL QUEUING MODELS

The general type of mathematical/statistical model most commonly applied to queuing problems is the Markov chain. The reason for this is that virtually any physical or logistical process involving the movement of resources or information can be visualized as a ramified network. Such networks lend themselves to investigation by Markov analysis. All of the following models are based upon a special Markov process known as the "birth-death" process. They each assume that inputs are generated by a Poisson process and that service times are described by the negative exponential distribution. The models will follow the notation of Budnick, Mojena and Vollman[1].

The Basic Models

Assume $\lambda_n = \lambda$ for all n and let $\rho = \frac{\lambda}{c\mu}$ be the system utilization factor. Then, for the case where c = 1 (the single queue-single server M/M/1 model using Kendall's notation), we have:

$$\left.\begin{aligned} \lambda_n &= \lambda \\ \mu_n &= \mu \end{aligned}\right\} \quad \text{for all n}$$

If $\rho < 1$, then this model will eventually attain steady state. Some important results for this model are:

$$P_o = 1 - \rho \tag{5}$$

$$P(n > 0) = \rho \tag{6}$$

$$P_n = P_o \rho^n \tag{7}$$

$$L_s = \frac{\lambda}{\mu - \lambda} \tag{8}$$

$$L_q = \frac{\rho\lambda}{\mu - \lambda} \tag{9}$$

$$L_b = \frac{\lambda}{\mu - \lambda} \tag{10}$$

$$W_s = \frac{1}{\mu - \lambda} \tag{11}$$

$$W_q = \frac{\rho}{\mu - \lambda} \tag{12}$$

$$W_b = \frac{1}{\mu - \lambda} \tag{13}$$

For the case where $c > 1$ (the single queue - multiple server M/M/c model from Kendall's notation) we have:

$$\lambda_n = \lambda \quad \text{for all } n$$

$$\mu_n = \begin{cases} n\mu \text{ for } n \leq c \\ c\mu \text{ for } n \geq c \end{cases}$$

The steady state requirements are $p = \frac{\lambda}{c\mu} < 1$ and the model results are:

$$P_o = \frac{1}{\left(\sum_{i=o}^{c-1} \frac{\rho^i}{i!}\right) + \frac{\rho^c}{c!\left(1 - \frac{\rho}{c}\right)}} \tag{14}$$

$$P_n = \begin{cases} \frac{\rho^n}{n!} \cdot P_o & \text{for } 0 \leq n \leq c \\ \frac{\rho^n}{c!c^{n-c}} \cdot P_o & \text{for } n \geq c \end{cases} \tag{15}$$

---

[1]F. S. Budnick, R. Mojena, and T. E. Vollmann, Principles of Operations Research for Management (Homewood, Il: Richard D. Irwin, Inc., 1977), Chap 12.

$$P(n \geq c) = \frac{\rho^c \mu c}{c!(\mu c-\lambda)} \cdot P_o \tag{16}$$

$$L_q = \frac{\rho^{c+1} \cdot P_o}{(c-1)!(c-\rho)^2} \tag{17}$$

$$L_s = L_q + \rho \tag{18}$$

$$L_b = \frac{L_q}{P(n \geq c)} \tag{19}$$

$$W_s = \frac{L_q}{\lambda} + \frac{1}{\mu} \tag{20}$$

$$W_q = \frac{L_q}{\lambda} \tag{21}$$

$$W_b = \frac{W_q}{P(n \geq c)} \tag{22}$$

Finite Queues

If the number of units in the system cannot exceed some number, say N, then the models must be modified accordingly. In these cases we have:

$$\lambda_n = \begin{cases} \lambda & \text{for } n = 0, 2, \ldots, N-1 \\ 0 & \text{for } n \geq N \end{cases}$$

$$\mu_n = \mu \quad \text{for all } n$$

There are no steady state requirements since $\lambda_n = 0$ for some state.

For the c = 1 case (M/M/1 model), the results are:

$$P_o = \begin{cases} \dfrac{1-\rho}{1-\rho^{N+1}} & \text{for } \lambda \neq \mu \\[2ex] \dfrac{1}{N+1} & \text{for } \lambda = \mu \end{cases} \tag{23}$$

$$P(n > 0) = 1-P_o \tag{24}$$

$$P_n = P_o\rho^n \quad \text{for } n \leq N \tag{25}$$

$$L_s = \begin{cases} \dfrac{\rho}{1-\rho} - \dfrac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} & \text{for } \lambda \neq \mu \\ \dfrac{N}{2} & \text{for } \lambda = \mu \end{cases} \tag{26}$$

$$L_q = L_s - (1-P_o) \tag{27}$$

$$L_b = \frac{L_q}{1-P_o} \tag{28}$$

$$W_s = \frac{L_q}{\lambda(1-P_N)} + \frac{1}{\mu} \tag{29}$$

$$W_q = W_s - \frac{1}{\mu} \tag{30}$$

$$W_b = \frac{W_q}{1-P_o} \tag{31}$$

For the cases where $c > 1$ (M/M/c model), the following results are offered (asume $c \leq N$):

$$P_o = \frac{1}{\left(\sum_{i=0}^{c} \frac{\rho^i}{i!}\right) + \left(\frac{1}{C!}\right) \cdot \left(\sum_{i=c+1}^{N} \frac{\rho^i}{c^{i-c}}\right)} \tag{32}$$

$$P_n = \begin{cases} \dfrac{\rho^n}{n!} \cdot P_o & \text{for } 0 \leq n \leq c \\ \dfrac{\rho^n}{c!c^{n-c}} \cdot P_o & \text{for } c \leq n \leq N \end{cases} \tag{33}$$

$$P(n \geq c) = 1 - P_o \sum_{i=0}^{c-1} \frac{\rho^i}{i!} \tag{34}$$

$$L_s = \frac{P_o \rho^{c+1}}{(c-1)!(c-\rho)^2} \left[1 - \left(\frac{\rho}{c}\right)^{N-c} - (N-c)\left(\frac{\rho}{c}\right)^{N-c} \cdot \left(1 - \frac{\rho}{c}\right)\right] + \rho(1-P_N) \tag{35}$$

$$L_q = L_s - \rho(1 - P_N) \tag{36}$$

$$L_b = \frac{L_q}{P(n \geq c)} \tag{37}$$

$$W_s = \frac{L_q}{\lambda(1-P_N)} + \frac{1}{\mu} \tag{38}$$

$$W_q = W_s - \frac{1}{\mu} \tag{39}$$

$$W_b = \frac{W_q}{P(n \geq c)} \tag{40}$$

Finite Populations

If the size of the calling population is finite, say of size m, then λ is now defined as the effective arrival rate $\lambda(m-L_s)$, $\mu_n = \mu$ for all n, and the models are modified as follows:

For the case where c = 1:

$$P_o = \frac{1}{\sum_{i=0}^{m} \left[\frac{m!}{(m-i)!} \cdot \rho^i \right]} \tag{41}$$

$$P(n > 0) = 1 - P_o \tag{42}$$

$$P_n = \frac{m!}{(m-n)!} \rho^n P_o \qquad \text{for } n \leq m \tag{43}$$

$$L_s = m - \frac{1}{\rho}(1-P_o) \tag{44}$$

$$L_q = m - \frac{(\lambda+\mu)(1-P_o)}{\lambda} \tag{45}$$

$$L_b = \frac{L_q}{1-P_o} \tag{46}$$

$$W_s = \frac{m}{\mu(1-P_o)} - \frac{1}{\lambda} \tag{47}$$

$$W_q = \frac{1}{\mu}\left(\frac{m}{1-P_o} - \frac{\lambda+\mu}{\lambda}\right) \tag{48}$$

$$W_b = \frac{W_q}{1-P_o} \tag{49}$$

For the case where $c > 1$:

$$P_o = \frac{1}{\left(\sum_{i=0}^{c} \frac{m!}{(m-i)!i!} \cdot \rho^i\right) + \left(\sum_{i=c+1}^{m} \frac{m!}{(m-i)!c!c^{i-c}} \cdot \rho^i\right)} \tag{50}$$

$$P_n = \begin{cases} \dfrac{m!P_o\rho^n}{(m-n)!n!} & \text{for } 0 \leq n \leq c \\[2ex] \dfrac{m!P_o\rho^n}{(m-n)!c!c^{n-c}} & \text{for } c \leq n \leq m \end{cases} \tag{51}$$

$$P(n \geq c) = 1 - P_o \sum_{i=0}^{c-1} \frac{m!}{(m-i)!i!} \cdot \rho^i \tag{52}$$

$$L_s = \frac{L_q + m\rho}{1+\rho} \tag{53}$$

$$L_q = \sum_{n=c+1}^{m} (n-c)P_n \tag{54}$$

$$L_b = \frac{L_q}{P(n \geq c)} \tag{55}$$

$$W_s = \frac{L_s}{\lambda(m-L_s)} \tag{56}$$

$$W_q = \frac{L_q}{\lambda(m-L_s)} \tag{57}$$

$$W_b = \frac{W_q}{P(n \geq c)} \tag{58}$$

Final Note

It should be obvious from the complexity of the formulas presented for these simple queuing situations that the applicability of mathematical techniques is limited for many practical problems. When such customer behaviors as balking, reneging, jockeying and colluding are considered along with complex structural networks of interacting queues in a system, analytical techniques cannot even be derived.

Large scale, complex, dynamic systems can be modeled effectively, however, through the use of simulation. During the past twenty-five years, several programming languages have been developed which greatly enhance our ability to model and analyze complex systems. Some examples of these are Dynamo, GPSS, GASP IV, Q-GERT, and SLAM.

EXAMPLES OF MATHEMATICAL MODELS

Single Server Queuing Model with Poisson Arrivals and Exponential Service

A base supply officer is considering the establishment of a customer relations desk at base supply to assist customers with supply related problems. The supply officer estimates that customers will arrive for assistance at the rate of 15 per hour. The supply clerk being considered for the customer relations job can service customers at the rate of one every three minutes. The supply officer wishes to determine the following:

a. Utilization of the supply clerk.

b. Average number of customers in the waiting line for assistance.

c. Average number of customers either waiting for assistance or in the process of receiving assistance.

d. Average waiting time in the line.

e. Average time spent waiting in line and receiving assistance.

f. The service rate necessary to reduce the average number of customers in the total system to two or less customers.

Assumptions/Properties of the Model:

1. Single server layout with a single queue.

2. Single service phase.

3. Infinite source population.

4. Poisson arrival pattern.

5. First come, first served queue discipline.

6. Exponential service pattern.

7. Unlimited permissible queue length.

Solution:

a. The average utilization of the supply clerk (ρ) is given by:

$$\rho = \frac{\lambda}{\mu} \text{ where,}$$

λ = customer arrival rate (arrivals per hour),

μ = customer service rate (services per hour).

The stated service rate of one customer every three minutes must be converted to the proper terms (services per hour).

$$\mu = \frac{1 \text{ service}}{3 \text{ minutes}} \times \frac{60 \text{ minutes}}{1 \text{ hour}} = 20 \text{ services per hour.}$$

Then, by substitution:

$$\rho = \frac{15}{20} = .75 = 75 \text{ percent utilization for the clerk.}$$

b. The average number in the waiting line (L_q) is given by substitutions into equation (9):

$$L_q = \frac{.75(15)}{20-15} = 2.25 \text{ customers}$$

c. The average number of customers in the waiting line or receiving assistance (average number of customers in the total system, L_s) is determined by equation (8):

$$L_s = \frac{15}{20-15} = 3 \text{ customers}$$

d. The average waiting time in the line (W_q) is given by equation (12):

$$W_q = \frac{.75}{20-15} = .15 \text{ hour or 9 minutes per customer}$$

e. The average waiting time in the line and receiving assistance (average waiting time in the total system, W_s) is given by equation (11):

$$W_s = \frac{1}{20-15} = .20 \text{ hour or 12 minutes per customer.}$$

f. The service rate necessary (for the supply clerk) to reduce the average number of customers in the total system to two or less is determined using equation (8):

$$L_s = \frac{15}{\mu-15} \leq 2,$$

$$2\mu - 30 \geq 15,$$

$$\mu \geq 22.5 \text{ services per hour}$$

Multiple Server Queuing Model with
Poisson Arrivals and Exponential Service

In the direct exchange section of a maintenance supply facility, activities are authorized to exchange malfunctioning assemblies and subassemblies for good ones. The bad components are then forwarded to the appropriate maintenance shops for repair. Normally, a mechanic from the using activity will present the bad assembly to the clerk in the direct exchange section. The clerk fills out the required paper work and exchanges the item while the mechanic waits. Mechanics arrive in a random (Poisson) pattern at the rate of 40 per hour and a clerk can fill requests at the rate of 20 per hour (exponential]. If the cost for a supply clerk is $2.00 per hour and the cost for a mechanic is $4.50 per hour, determine the optimum number of clerks to staff the direct exchange section.

Assumptions/Properties of the Model:

1. Multiple server layout with a single queue.

2. Single service phase.

3. Infinite source population.

4. Poisson arrival pattern.

5. First come, first served queue discipline.

6. Exponential service pattern.

7. Unlimited permissible queue length.

Solution:

The first step is to determine the average number of mechanics waiting in line at the direct exchange counter. The cost of this "idle" time is then compared to the cost of adding additional clerks. To begin, assume that three clerks (c = 3) will be utilized because with a high arrival rate in relation to the service rate, a system with one or two servers would

result in excessively long lines. The average number of mechanics in the line is computed by equations (14) and (17) (note that $\rho = \lambda/\mu = 40/20 = 2$):

$$P_o = \frac{1}{\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!}\right) + \frac{2^3}{3!\left(1 - \frac{2}{3}\right)}} = .111$$

$$L_q = \frac{2^4\ (.111)}{(3-1)!\,(3-2)^2} = .888 \text{ mechanics}$$

For an eight hour day at \$4.50 per hour, the value of the mechanic's idle time can then be computed as: .888 mechanic x \$4.50 per hour x 8 hours = \$31.97.

The next step is to recalculate the waiting time if another clerk is added to the system. Then, compare the added cost of the additional clerk to the value of the time saved by the mechanics in the line. Using equations (14) and (17), when c = 4.

$$P_o = \frac{1}{\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!}\right) + \frac{2^4}{4!\left(1 - \frac{2}{4}\right)}} = .130$$

$$L_q = \frac{2^5\ (.130)}{(4-1)!\,(4-2)^2} = .173 \text{ mechanics}$$

The value of the mechanic's idle time with four clerks is as follows: .173 mechanic x \$4.50 per hour x 8 hours = \$6.23. Therefore, the value of mechanic's time saved in the line is \$31.97 - \$6.23 = \$25.74. The cost of the additional clerk is \$2.00 per hour x 8 hours = \$16.00. The cost reduction obtained by adding the fourth clerk is \$25.74 - \$16.00 = \$9.74, and the addition of a fourth clerk is thus advantageous.

The next step is to recalculate the waiting time if a fifth supply clerk is added and compare the additional cost with the value of the time saved by the mechanics in line. From equations (14) and (17), it can be determined the L_q = .0397 mechanics. The cost of idle time is now .0397 mechanics x \$4.50 per hour x 8 hours = \$1.43. The value of mechanic's time saved (compared to the three clerk system) is \$31.97 - \$1.43 = \$30.54. The cost of two additional clerks is 2 clerks x \$2.00 per hour x 8 hours = \$32.00, which is greater than the value of the time saved. These results, therefore, suggest that four clerks should be used in this multiple server system.

Singel Server Queuing Model with Limited Queue Length

A NATO country is requesting USAF airlift support for a large scale movement requirement of recently procured explosive cargo. It is proposed that the airlift aircraft, to be chartered under AFR 76-28 rate tariffs, will offload the munitions at a USAF installation which is colocated on a civilian airfield within the NATO country. The installation's aerial port squadron plans to use an isolated offload ramp, which has space to offload one aircraft at a time, and an entrance taxiway which has space for three aircraft waiting to be offloaded. The average arrival rate of airlift aircraft is estimated at two aircraft per hour and the offload rate is estimated at 2.5 aircraft per hour. The marginal revenue derived from the rate tariffs to cover certain operating expenses, is estimated at $500 per aircraft load. Additional ramp space for holding aircraft waiting to be offloaded is available on the civilian airfield at a leased rate of $250 per day per space. The airfield will be open 14 hours a day for this operation.

Assuming Poisson arrivals and exponential service, should additional aircraft parking space be rented on the civilian airfield to expand the restricted queue? If so, how many spaces? In addition, what will be the average number of aircraft in the queuing system, both in line and being offloaded?

Assumptions/Properties of the Model:

1. Single server layout with a single queue.

2. Single service phase.

3. Infinite source population.

4. Poisson arrival pattern.

5. First come, first served queue discipline.

6. Exponential service pattern.

7. Limited permissible queue length.

Solution:

One approach to solving this problem is to assume an increasing number of aircraft parking spaces and compare the additional revenue generated by each additional space. Additional spaces will be rented until the revenue becomes less than the $250 cost of leasing.

Aircraft will be offloaded at the rate of 2.5 aircraft per hour any time there are aircraft in the queuing system. To find the amount of time there will be aircraft in the system, we can solve for the probability of zero in the system using quation (23) and subtract this result from 1.00 in equation (24) to derive the percent of time that aircraft are being offloaded (alternatively, the percent of time that the system is busy).

For the probability of zero aircraft in the system with four spaces available in the total queuing system:

$$\rho = \frac{\lambda}{\mu} = \frac{2}{2.5} = .8$$

$$n = 0$$

$$N = 4$$

$$P_o = \frac{1-.8}{1-.8^{4+1}} = .297$$

$$P(n > 0) = 1 - P_o = 1 - .297 = .703$$

Therefore, aircraft will be in the process of being offloaded (the system will be busy) 70.3% of the time.

When one space is rented, N = 4 +1, and equations (23) and (24) give:

$$P_o = \frac{1-.8}{1-.8^{5+1}} = .271$$

$$P(n > 0) = 1 - .271 = .729$$

In this case, service will be carried out 72.9% of the time, or an increase of 2.6% (72.9% - 70.3%). In revenue, this is worth .026 x 2.5 aircraft per hour x 14 hours x $500 per aircraft or $455. This is considerably more than the cost of leasing the additional space.

When a second parking space is leased, N = 4 + 2, and equations (23) and (24) give:

$$P_o = \frac{1-.8}{1-.8^{6+1}} = .253$$

$$P(n > 0) = 1 - .253 = .747$$

Thus, with six total spaces in the queuing system, the aerial port squadron will be busy 74.7% of the time offloading aircraft. The additional space increases utilization of the aerial port squadron by 1.8% x 2.5 aircraft per hour x 14 hours x $500 per aircraft, or $315. Since this amount is greater than the $250 rental charge, the space should be rented.

A third leased space, with N = 4 + 3, gives:

$$P_o = \frac{1-.8}{1-.8^{7+1}} = .240$$

$$P(n>0) = 1 - .240 = .760$$

The increased service utilization is 1.3% (76.0% - 74.7%) and the added revenue is given by: .013 x 2.5 aircraft per hour x 14 hours x $500 per aircraft = $227.50. Thus, the third space should not be rented since it would incur a loss of $22.50 ($250 - $227.5).

A reasonable solution, then, is to lease just two additional spaces. The revenue is given by: .747 x 2.5 aircraft per hour x 14 hours x $500 per aircraft or $13,072.50.

To determine the average number of aircraft in the total system with two additional spaces to be leased, we use equation (26):

$$N = 6$$

$$L_s = \frac{.8}{1-.8} - \frac{(6+1).8^{6+1}}{1-.8^{6+1}} = 2.14 \text{ aircraft.}$$

REFERENCES

Budnick, Frank S., Richard Mojena and Thomas E. Vollmann. Principles of Operations Research for Management. Homewood, Il.: Richard D. Irwin, Inc. 1977.

Chase, Richard B. and Nicholas J. Aquilano. Production and Operations Management. Homewood, Il.: Richard D. Irwin, Inc., 1977.

Hillier, Frederick S. and Gerald J. Lieberman. Introduction to Operations Research. San Francisco, Ca.: Holder-Day, Inc., 1967.

Wagner, Harvey M. Principles of Operations Research. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1975.

CHAPTER 20

RELIABILITY

INTRODUCTION

The logistics manager is constantly faced with the problem of increasing support costs in an era of scarce resources. Numerous studies have shown that, in large part, support costs are related to system and component reliability. For this reason it becomes increasingly important for the logistics manager to have an understanding of the basic concepts and techniques of reliability theory. The following sections present some of the tools which can be applied by the logistics manager to problems and decisions in the area of reliability. In no way is this chapter intended to be a complete treatise on reliability theory. The models are presented without derivation or proof. Numerous examples are provided to illustrate the application of the principles and models in the logistics arena.

DEFINITIONS

Numerous terms and phrases are used consistently in the literature and in practice. These are listed below along with their accepted usage.

1. Availability - A measure usually stated as a probability that a system will operate satisfactorily when called upon to operate under stated conditions, usually in an operational environment. Mathematically, availability is given as:

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}$$

2. Failure Rate - The rate at which failures occur during a specified interval of time, t to t + a, given by:

$$\lambda\ (t,a) = \frac{R(t) - R(t+a)}{aR(t)}$$

3. Hazard Rate - Instantaneous rate of failure given as:

$$z(t) = \lim_{a \to o} \lambda\ (t,a)$$

$$= \frac{f(t)}{R(t)}$$

where $f(t)$ = the failure rate distribution

and $R(t)$ = the reliability function which can be found by

$$R(t) = 1 - \int_0^t f(t)dt$$

$$= 1 - F(t)$$

4. Mean Time Between Failures (MTBF) - The average life of a system or equipment. MTBF is considered to be a primary measure of reliability. $MTBF = \int_0^\infty R(t)dt$.

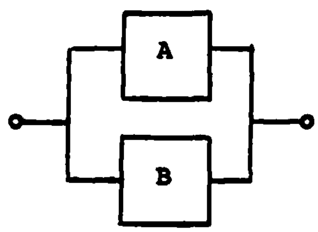
5. Mean Time to Failure (MTTF) - Means the same as MTBF except that MTTF is used for items which will not be repaired.

6. Mean Time to Repair (MTTR) - The average time to repair an item including both scheduled and unscheduled maintenance.

7. Median Life - The time period for which the system/equipment reliability is 0.50, i.e., half of the items will fail after their median life and half before the median life.

8. Parallel Design - Two or more components are connected in such a way that each must fail in order for the entire system to fail.

Figure 20-1



Since this system has alternate paths of operation, both A and B must fail for the entire system to fail.

9. Redundancy - Associated with parallel design. Redundancy usually increases reliability by providing alternate paths of operation.

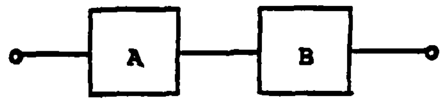
10. Reliability - The probability that a system or equipment will operate over a stated period of time for its intended purpose under stated conditions.

11. Reliability Growth - Improvement (increase) in an item's reliability after the "burn-in" period due to the conscious efforts of engineers and management usually requiring an expenditure of resources.

12. Reliability Improvement Warranty (RIW) - A contractual technique whereby the contractor is provided with incentives to reduce system support costs through reliability improvements over a stated period of time.

13. Series Design - System/equipment design configuration in which all components must operate for the system to operate.

Figure 20-2

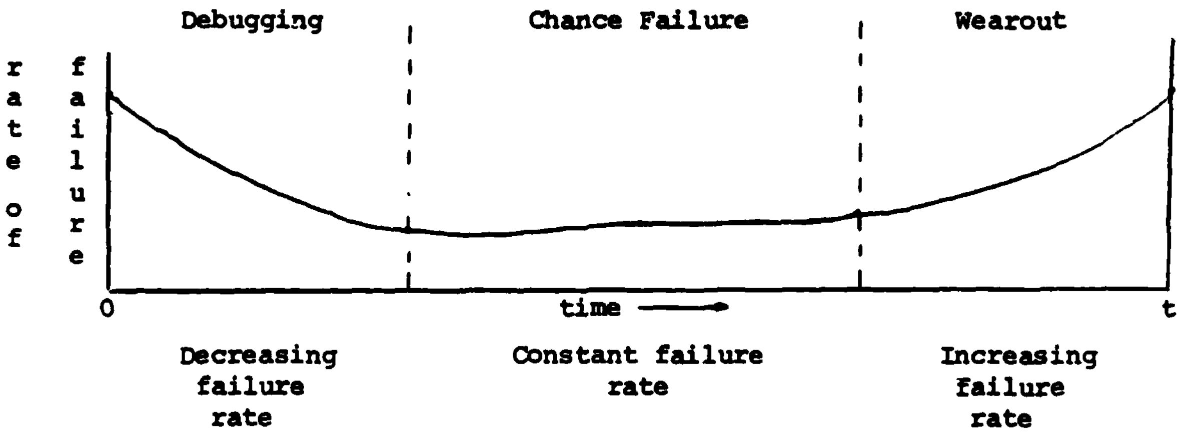


If component A or B fails, the system fails.

BASIC CONCEPTS

Reliability is concerned with system/component failure. These failures have been categorized into three types which are illustrated in the figure below.

Figure 20-3



1. Early failure - occurs during system "burn-in" or debugging phase These are a result of the manufacturing or assembling process when the equipment and/or techniques are new or highly complex.

2. Wearout failure - Usually a result of the aging process and when items are not properly maintained or parts replaced when required.

3. Chance failure - Occurs randomly throughout the item's useful life and is the most difficult to reduce or eliminate.

DETERMINING RELIABILITY

The theory and practice of reliability is closely associated with the application of statistical probability. In reliability, only two outcomes are possible--success or failure. Either a system will operate,

or it won't. These outcomes are then mutually exclusive since both cannot occur at any one point in time. It is normally the best policy to define failure and let all other possibilities be defined as success.

After the exact definition of failure has been specified, the frequency at which failures occur then becomes a parameter for the statistical formulation of reliability. This parameter is called a failure rate which is expressed as the number of failures per period of time. Its reciprocal is the mean time between failures (MTBF) defined in the previous section.

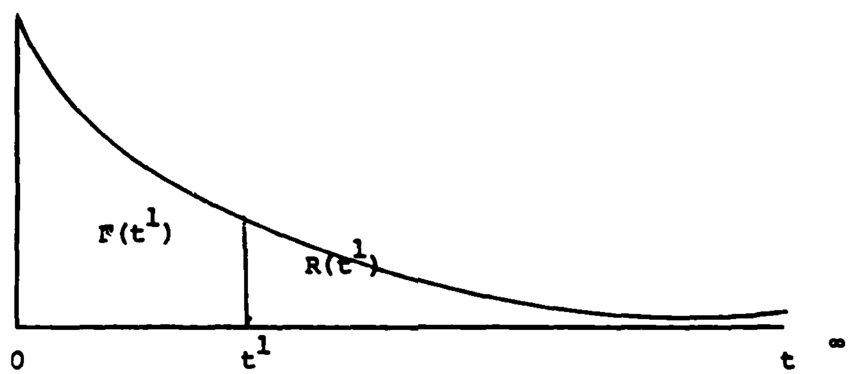
Failure rates are determined experimentally by testing the systems under study or by observing their performance over a period of time. It is usually best to obtain a large number of trials before establishing firm statements about reliability. For example, we would not hurriedly assume that a coin is biased if three heads and seven tails appeared in ten tosses. In this respect, we should not automatically jump to conclusions about a system's reliability based on a relatively short experiential period.

In determining reliability, probability theory plays an important role. Analytically, the reliability of a system at a given point in time, t, is expressed as:

$$R(t^1) = 1 - F(t^1)$$

The interpretation is in probabilistic terms. Thus, $R(t^1)$ is the probability that a failure occurs after time t^1. Conversely, $R(t^1)$ is the probability that the system will not fail before t^1. Graphically, this principle idea is shown below.

Figure 20-4



The amount of time to failure is defined as a probability density function, f(t), where

$$f(t) = \frac{dF(t)}{dt}$$

Without proof, it can be stated that reliability, R(t), is represented by an area under the time to failure density curve, i.e., a probability.

Therefore, $0 \leq R(t) \leq 1$

Once the time to failure distribution is known, i.e., its probability density function, the reliability of the system can be readily determined.

$$\text{Since} \quad f(t) = \frac{dF(t)}{dt}$$

$$\text{and} \quad R(t) = 1 - F(t)$$

$$\text{then} \quad R(t) = 1 - \int_0^t f(t)dt$$

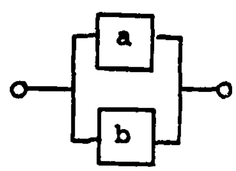
In other words, the reliability of a system or the probability that it will operate successfully up to time t is found by integrating the density function from o to t and subtracting the result from one.

Obviously, finding the reliability of today's technologically complex systems is not so straightforward and simple. Yet, this basic principle applies. Let us proceed to demonstrate the calculation of reliability under several different circumstances.

When system design and the reliability of components are known, system reliability can be obtained by directly applying the laws of probability and Boolean algebra.

Figure 20-5

Components in Parallel



R_a (reliability of component a) $= R_b = 0.7$

$$R(S) = R_a \cup R_b = \text{reliability of the system}$$

$$= R_a + R_b - R_a \cdot R_b$$

$$= 0.7 + 0.7 - [(0.7)(0.7)]$$

$$= 1.4 - .49 = 0.91$$

For components in parallel, the probability rule for the union of events applies. Further, in all cases, the reliability for each component is taken to be independent of all others.

Figure 20-6

Components in Series

$$R_a = R_b = 0.7$$

$$R(S) = R_a \cap R_b = R_a \cdot R_b$$

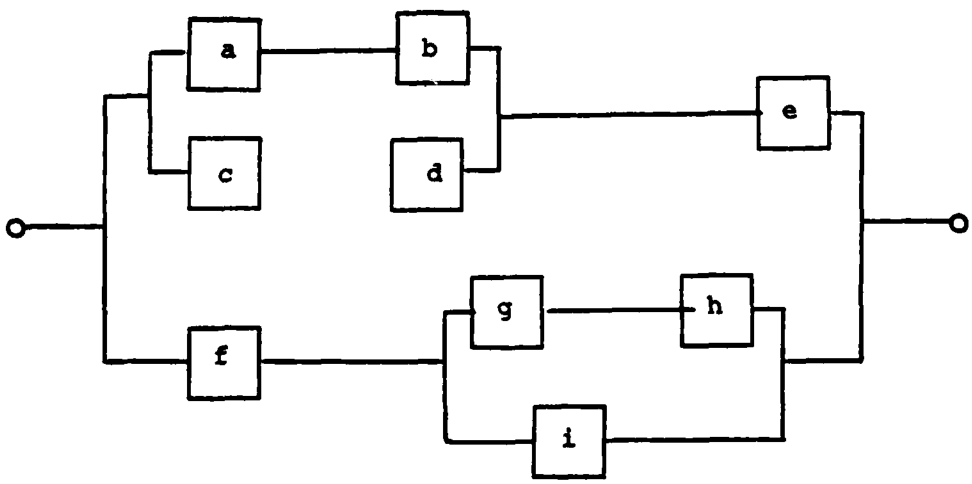
$$= (0.7)(0.7)$$

$$= 0.49$$

The multiplication rule applies since a and b must operate or the system fails. This implies that no matter how high component reliability may be, system reliability decreases as more components are added in series.

The following example illustrates a combination of parallel and series designs, but uses the same probability rules in determining system reliability.

Figure 20-7



The reliability of each component is:

$R_a = .92$ $\quad R_f = .96$

$R_b = .97$ $\quad R_g = .93$

$R_c = .98$ $\quad R_h = .98$

$R_d = .98$ $\quad R_i = .93$

$R_e = .98$

To find the reliability of the system, the following expression is solved: (Capital letters are used for the components)

$$R(S) = \{[(A \cap B) \cup (C \cap D)] \cap E\} \cup \{F \cap [(G \cap H) \cup I]\}$$

To simplify the problem, additional components are defined and substituted into the original expression.

$$\text{Let } R = (A \cap B) = (.92)(.97) = .8924$$
$$S = (C \cap D) = (.98)^2 = .9604$$
$$T = (G \cap H) = (.93)(.98) = .9114$$

The original expression becomes:

$$R(S) = [(R \cup S) \cap E] \cup [F \cap (T \cup I)]$$
$$= [(R + S - R \cdot S) \cdot E] \cup [F \cdot (T + I - T \cdot I)]$$
$$= [(.8924 + .9604 - .8924 \cdot .9604)(.98)] \cup$$
$$[.96 \cdot (.9114 + .93 - .9114 \cdot .93)]$$

Additionally, if we let

$$X = [(R + S - R \cdot S) \cdot E]$$
$$Y = [F \cdot (T + I - T \cdot I)]$$

we need only to solve for

$$R(S) = X \cup Y$$

$$= X + Y - X \cdot Y$$

$$= .97582 + .95404 - .97582 \cdot .95404 = .99888$$

System or component reliability is also determined through the use of reliability models. Three basic models are called: constant hazard, linear increasing hazard, and Weibull.

Constant Hazard Model. The constant hazard model has already been introduced in Chapter 7 as the exponential distribution, sometimes referred to as the negative exponential. It is characterized by a curve which is downward sloping to the right. This model is often used to determine the reliability of components, particularly electronic components, after the burn-in period, but prior to the wearout phase of the life cycle. During this period, failures are considered to occur randomly (by chance) rather than as the result of a specific failure mechanism. Within this period of operating life, the reliability (or probability of survival) is the same for all periods of equal length. The number of failures occurring during any time interval appears to be related only to the total number of units in operation and the period of time in the interval. For each period, the operating time begins at t = 0. The component or equipment is considered "new" at this time in that it has survived previous missions or operating periods.

Parameters	λ = failures per time period
Hazard function:	$z(t) = \lambda$
Density function of time to failure:	$f(t) = \lambda e^{-\lambda t}$
Cumulative distribution function of time to failure:	$F(t) = 1 - e^{-\lambda t}$
Reliability function:	$R(t) = e^{-\lambda t}$
Mean time to failure or mean time between failures:	$MTTR = \frac{1}{\lambda}$

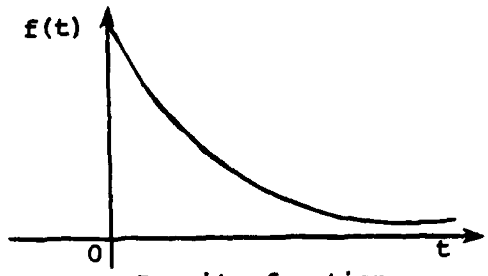
The figures below depict the exponential hazard and density functions.

Figure 20-8

Constant hazard

Note that the rate of failure, λ, is the same for all t.

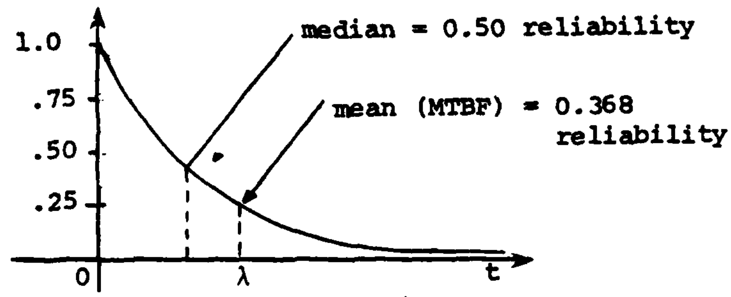
Figure 20-9



Density function

It should be noted that when the MTBF (or MTTF) is equal to t, the reliability is 0.368 as shown below.

Figure 20-10



Reliability function

It should also be noted that given a stated reliability over a time interval, mean life (MTBF or MTTR) can be easily found by:

$$\text{MTBF}\ \left.\frac{1}{\lambda}\right) = \frac{-t}{\log_e R(t)}$$

For example, if the reliability of a component for a 10-hour mission is .904, what is the component's mean life?

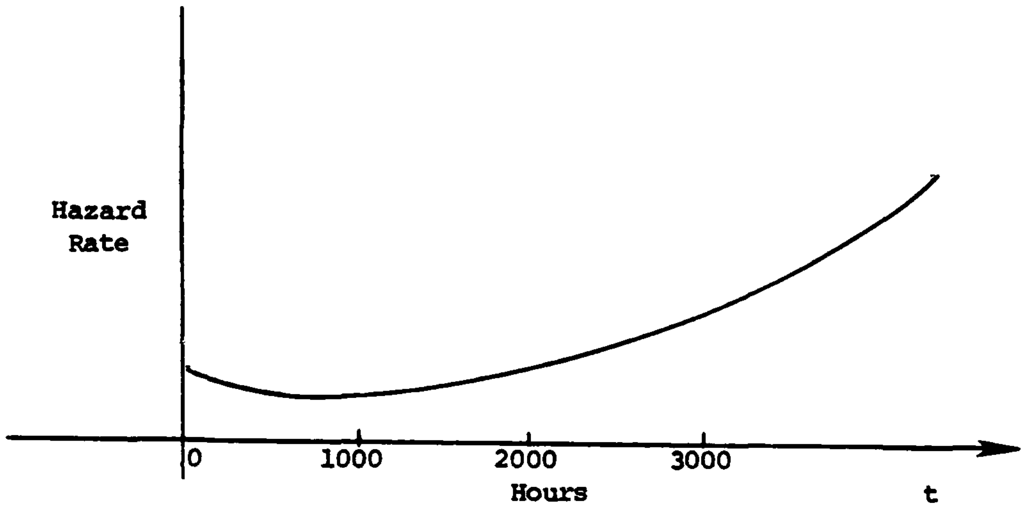
$$\text{MTBF} = \frac{-t}{\log_e R(t)}$$

$$\text{MTBF} = \frac{-10}{\log_e 0.904}$$

$$\text{MTBF} = 99.08 \text{ hours}$$

At this point a work of caution needs to be stated. It is important that the distribution of failures be known prior to making any statements about system/component reliability. In too many cases, the analyst elects to apply the constant hazard model because it is very easy to use. This may lead to incorrect assumptions about reliability which can result in failed missions or excessive costs for system support, or both. In this respect, many components/systems have been noted to have failure distributions which have changed over time as shown in the figure below.

Figure 20-11



In this case, a constant hazard model could be used for the first 2000 hours of operation after which a linear increasing hazard model might be applicable.

When the individual is quite sure that an exponential model is applicable, reliability can be determined from raw data from a test as follows.

<u>EXAMPLE 20-1</u>

Twenty items are put on test. Failures occurred at the following times given in hours: 20, 27, 30, 41, 42, 48, 60, 66, 80, 81, 83, 86, 105, 108, and 130. Five items survived past 130 hours.

First, compute an estimate of the rate of failure by:

$$\text{MTBF} = \frac{1}{\hat{\lambda}} = \frac{\text{total time for failed items} + \text{time for non-failed items}}{\text{number of failures}}$$

$$= \frac{1007 + 650}{15}$$

$$= 110.47 \text{ hours}$$

$$\text{MTBF} = \frac{1}{\hat{\lambda}}$$

$$\hat{\lambda} = \frac{1}{\text{MTBF}}$$

$$= \frac{1}{110.47}$$

$$= 0.00905$$

The reliability function is:

$$R(t) = e^{-\lambda t}$$

$$R(t) = e^{-\frac{t}{110.47}} = e^{-0.00905t}$$

The reliability of the item for a 10-hour mission, then is:

$$R(10) = e^{-\frac{10}{110.47}}$$

$$= 0.9135$$

Linear Increasing Hazard Model. The distribution for this model is called the Rayleigh distribution. This distribution is usually most applicable when systems/components begin to wear out.

Parameter: α = slope of hazard curve

Hazard function: $z(t) = \alpha t$

Density function of time to failure $\alpha t e^{-\frac{\alpha t^2}{2}}$

Cumulative distribution function of time to failure: $F(t) = e^{-\frac{\alpha t^2}{2}}$

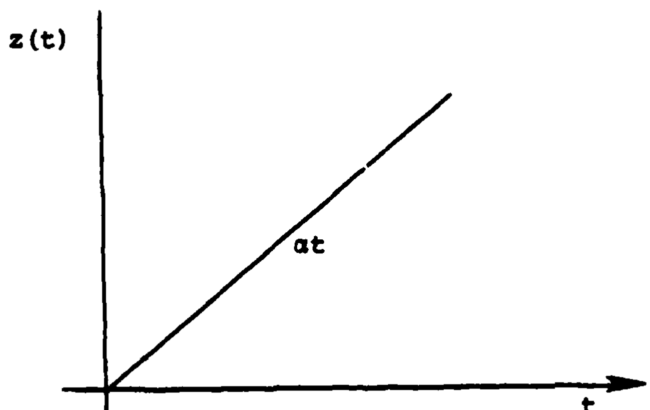
Reliability function: $R(t) = 1 - F(t)$

$$= e^{-\frac{\alpha t^2}{2}}$$

Mean time to failure: $MTTF = \sqrt{\frac{\pi}{(2\alpha)}}$

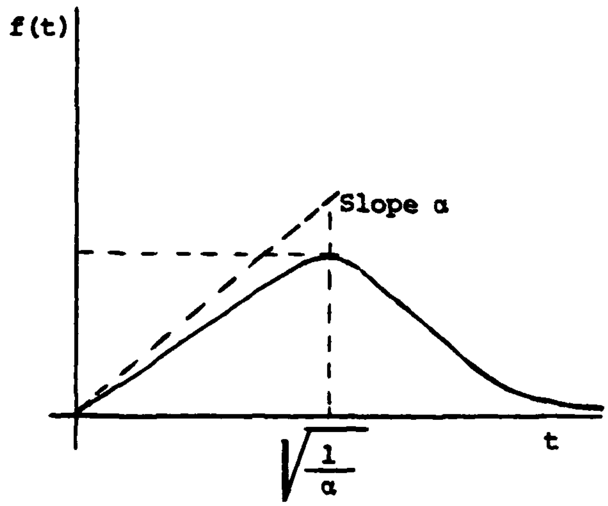
Graphical descriptions of the linear increasing hazard function and density function are given in figures 20-12 and 20-13.

Figure 20-12



Linear increasing hazard

Figure 20-13



Rayleigh density function

EXAMPLE 20-2

Failure data collected on the Guidance and Control Unit of a missile indicate that the unit failures follow a linear increasing model with $\alpha = .0022$. What is the unit's MTBF and reliability for a missi^n time of 12 hours?

$$\text{MTBF} = \sqrt{\frac{\pi}{2\alpha}}$$

$$= \sqrt{\frac{3.1416}{(2)(.0022)}}$$

$$= 26.72 \text{ hours}$$

To find the reliability for a 12-hour mission:

$$R(12) = e^{-\frac{(.0022)(12)^2}{2}}$$

$$= 0.8535$$

Weibull. The most complex of the distributions used in reliability, the Weibull distribution has great flexibility and is used widely in assessing failure distributions of semi-conductors, mechanical devices, and component unit levels.

Parameters: α = scaling parameter

β = shaping parameter

γ = location parameter

The Weibull is actually a family of distributions depending upon the values of its parameters. For example, when $\beta = 1$ and $\gamma = 0$, the distribution is exponential and α is the MTBF, i.e., $\frac{1}{\lambda}$ in the constant hazard model. At the same time for $\beta < 1$ the hazard rate is a decreasing function, and for $\beta > 1$ the hazard rate increases with time. When $\alpha = 1$ and $\gamma = 0$, the distribution is near normal when $\beta \approx 3.2$. In reliability applications, the Weibull is generally used with only two parameters α and β, with $\gamma = 0$. The values for the parameters can be obtained with reasonable accuracy by plotting failure data on Weibull probability paper. The reader should review more advanced texts in reliability theory for a complete explanation. The figures below illustrate various shapes of the Weibull with different values of α and β.

Hazard function: $$z(t) = \frac{\beta t^{\beta-1}}{\alpha}$$

Density function of time to failure: $$f(t) = \frac{\beta t^{\beta-1}}{\alpha} e^{\frac{-t^{\beta}}{\alpha}}$$

Cumulative distribution function $$F(t) = 1 - e^{\frac{-t^{\beta}}{\alpha}}$$

Reliability function: $R(t) = 1 - F(t)$ $$R(t) = e^{\frac{-t^{\beta}}{\alpha}}$$

Mean time to failure: $MTTF = \int_0^{\infty} R(t)\,dt$ $$MTTF = \alpha^{\frac{1}{\beta}} \Gamma\left(\frac{1}{\beta} + 1\right)$$

where $\Gamma(n + 1) = n\,\Gamma(n)$ *

or $\Gamma(n) = n-1\,\Gamma(n-1)$

*See Selby, S. M. (ed), CRC Standard Mathematical Tables. Cleveland, The Chemical Rubber Company, 1975.

Figure 20-14

Weibull distributions with Different Values of α; β = 2; γ = 0

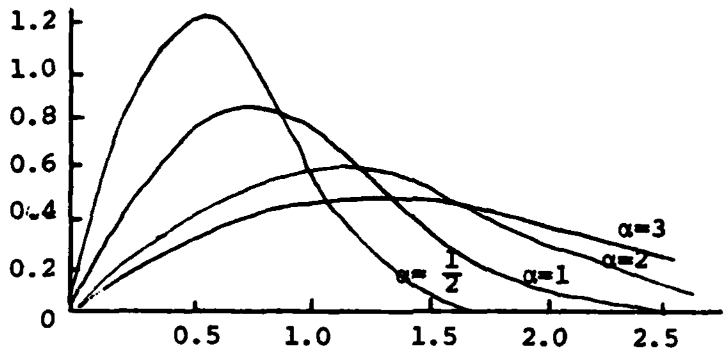
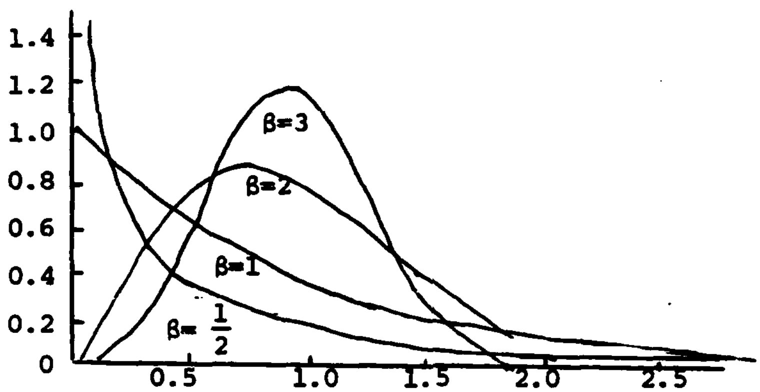


Figure 20-15

Weibull Distributions with Different Values of β; α = 1, γ = 0



The density and reliability functions are somewhat different when the third parameter γ is added. This location parameter is used only when there is a substantial time period during which no failures are likely to occur. For discussions of the three parameter Weibull distribution, texts on reliability mathematics should be consulted.

EXAMPLE 20-3

Failures of tires on a particular aircraft follow a distribution pattern which is closely approximated by a two-parameter Weibull model with $\alpha = 262.79$ and $\beta = 1.13$. This distribution was determined according to the following relationship: 2 take-offs + 2 landings = 1 "time unit." What is the mean time to failure for this type of tire? What is the reliability of the tire if its mission length is set at 14 take-offs and 14 landings?

$$\text{MTTF} = 262.79^{\frac{1}{1.13}} \; \Gamma\left(\frac{1}{1.13} + 1\right)$$

$$= 262.79^{.885} \; \Gamma\,(1.885)$$

$$= (138.469)(.95673)^{*}$$

$$= 132.48$$

The MTTF must be converted to "Time Units" which would be 2 x 132.48 $\approx$ 265 take-offs and landings.

*$\Gamma(1.885)$ interpolated from table on p. 533, CRC Standard Mathematical Tables, 1975.

To find the reliability of 14 take-offs and landings, the conversion would make this equivalent to $\frac{14}{2}$ or 7 time units.

$$R(t) = e^{-\frac{7^{1.13}}{262.79}}$$

$$= 0.9663$$

To illustrate the Weibull as a constant hazard model ($\beta = 1$), the previous example for the constant hazard model will be reiterated. Recall that λ in the example was found to be 0.00905. In the Weibull model, $\alpha = \lambda$.

$$\text{MTTF} = 110.497^{\frac{1}{1}} \, \Gamma\left(\frac{1}{1} + 1\right)$$

$$= (110.497)(1)$$

$$= 110.497 \text{ hours}$$

To find the reliability for a 10-hour mission:

$$R(t) = e^{-\frac{10^{1}}{110.497}}$$

$$= 0.9135$$

The similarity (and flexibility of the Weibull model) can thus be readily seen. The next section demonstrates a technique which uses raw failure data in identifying the best reliability model.

Identifying the model. Ten electronic components have been placed on test for a continuous period until all ten failed. The time of failure of each unit was recorded in Table 20-1.

Table 20-1

Failure Data

Failure	Total Operating Time
1	8
2	20
3	34
4	46
5	63
6	86
7	111
8	141
9	186
10	266

The problem is to find the reliability of the component. In order to do so, the model of best fit must be determined. A logical place to start is to plot the failure data to determine whether time to failure follows a known distribution. The density function of time to failure can be defined as:

$$f(t) = \frac{\dfrac{n(t_i) - n(t_i + \Delta t_i)}{N}}{\Delta t_i}$$

where,

$n(t_i)$ = number of surviving items at the beginning of interval t_i

N = total number of items placed on test

Δt_i = length of interval, i.e., amount of time elapsed since end of last interval

Thus, for the first failure

$$f(t) = \frac{\dfrac{[10 - 9]}{10}}{8} \cdot 100 = 1.25$$

To simplify plotting the function, 100-hour intervals for t have been established. For failure number two:

$$f(t) = \frac{\dfrac{[9 - 8]}{10}}{12} \cdot 100 = 0.84$$

Table 20-2 shows the values computed for each failure. The plot of f(t) is given in Figure 20-16.

Next, the hazard function is plotted as shown in Figure 20-17. The hazard function is defined as:

$$z(t) = \frac{\dfrac{[n(t_i) - n(t_i - \Delta t_i)]}{n(t_i)}}{\Delta t_i}$$

For failure 1:

$$z(t) = \frac{\frac{[10 - 9]}{10}}{8} \cdot 100 = 1.25$$

Note that t is also presented in 100-hour increments to simplify plotting.

For failure 2:

$$z(t) = \frac{\frac{[9 - 8]}{9}}{12} \cdot 100 = 0.93$$

The values computed for z(t) for all failures is given in Table 20-2.

By inspecting Figures 20-16 and 20-17, it can be seen that time to failure appears to approach a constant hazard model, i.e., exponential distribution. Reliability for this component can then be ascertained as described earlier for the constant hazard model. Reliability from the raw data can be computed directly for intervals at any of the failure times as follows:

$$R(t) = \frac{n(t_i)}{N}$$

where $n(t_i)$ = number of survivors up to and including the i^{th} failure

N = total number of items placed on test

Table 20-2

Failure i	Operating Time	t	f(t)	Z(t)
1	8	8	1.25	1.25
2	20	12	0.84	0.93
3	34	14	0.72	0.96
4	46	12	0.84	1.19
5	63	17	0.59	0.98
6	86	23	0.44	0.87
7	111	25	0.40	1.00
8	141	30	0.33	1.11
9	186	45	0.22	1.11
10	266	80	0.13	1.25

Figure 20-16

Time to Failure Density Function

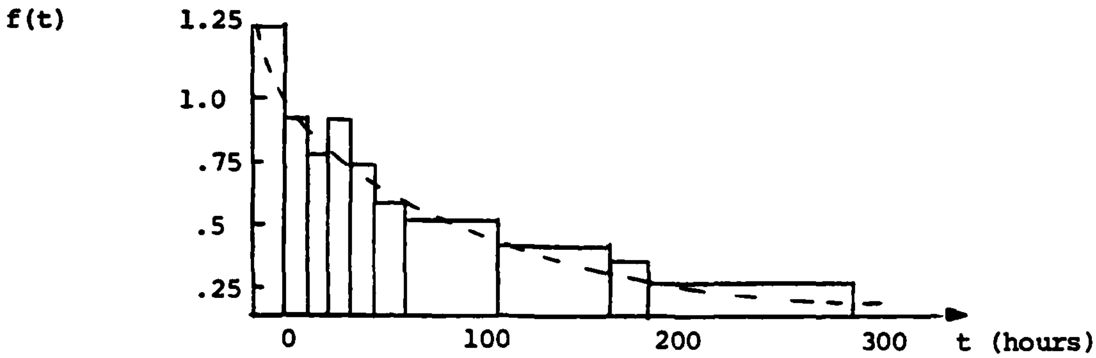
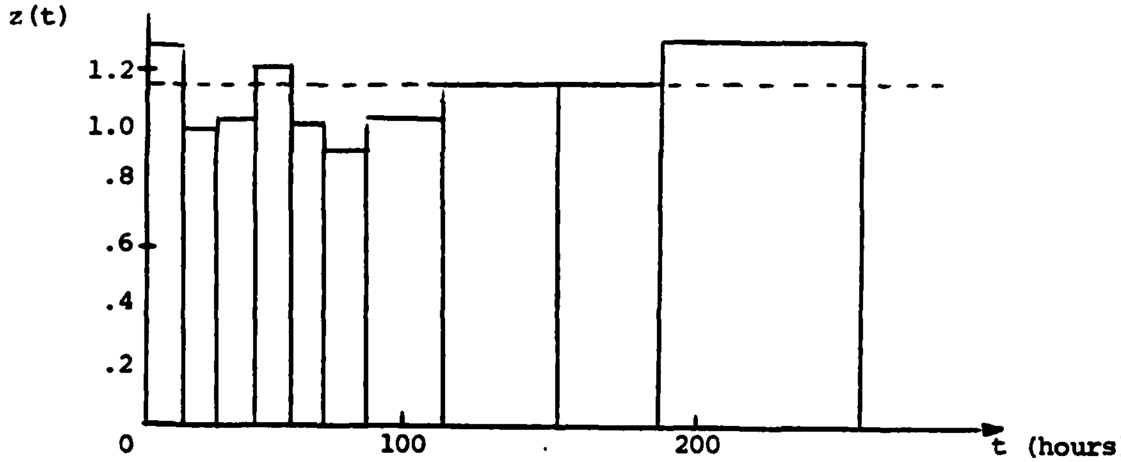


Figure 20-17

Hazard Function



For the fourth failure, t_i = 46 hours

and $R\left(t_{46}\right) = \frac{6}{10} = 0.60$

The reliability for a 46-hour period is 0.60.

At this point a note of extreme caution must be made. Reliability is a function of the number of failures, which makes the sample size of of paramount importance. When the sample size, i.e., the number of failures, is small, the analyst runs the same risk in predicting reliability that is usually encountered in other forms of statistical estimation. At the same time, the costs in dollars and time necessary to achieve an adequate amount of failure data, is almost always more than what is available The manager, theh, is forced to make a critical trade-off decision in this area.

Continuing with the example, when a large number of items is placed on test, the previous procedure may be cumbersome. An alternate method exists for using a class interval structure to analyze failure data for identifying an appropriate reliability model.

In Table 20-3, the number of hours for the testing period is divided into 1,000-hour intervals along with the number of failures occurring during each interval.

Table 20-3

Hours Δt_i	Number of Failures	f(t)	z(t)
0-1000	59	3.43	3.43
1001-2000	24	1.40	2.12
2001-3000	29	1.69	3.26
3001-4000	30	1.74	5.00
4001-5000	17	0.99	5.69
5001-6000	13	0.76	10.00
	172		

Again, the test is run until all items fail. The same formulation is used as with the individual failure items except that

t_i = time at the beginning of the class interval

Δt_i = length of the class interval

As usual with grouped data, failures are assumed to occur evenly throughout each class interval.

The failure density for class number 3 is found by:

$$f(t) = \frac{\dfrac{[89 - 60]}{172}}{1000} \cdot 10^4 = 1.69$$

The resulting density is multiplied by 10^4 to simplify graphing the data. The hazard function for the same class interval is found by:

$$z(t) = \frac{\frac{29}{89}}{1000} \cdot 10^4 = 3.26$$

When a large amount of failure data is available or planned, an appropriate class interval can be constructed using Sturges' rule.

$$k = 1 + 3.3 \log n$$

where k = number of classes rounded to nearest integer

n = total number of failures observed.

Although it is not necessary, it is usually best to make class intervals of equal width.

As a final note in this area, the use of grouped data has a disadvantage in that the exact time elapsed to each failure is lost within the class intervals. This loss of individual failure data items may affect the precision of the reliability estimates, but will usually not hinder the identification of the appropriate reliability model.

Curve Fitting. In many instances, inspecting a plot of time to failure data will not provide an identifiable probability distribution. Several curve-fitting techniques are available which will define a specific distribution for a given set of failure data. In addition to these curve-fitting techniques which are not discussed in detail here, statistical goodness-of-fit tests can be easily applied which give the manager reasonable assurance that a set of data follows a stated distribution. The Chi-square goodness-of-fit test as well as the Kolmogorov-Smirnov test is described in detail in Chapter 22.

SYSTEM RELIABILITY

As stated earlier, system reliability depends upon the reliability of each of the components as well as their configuration. A previous example has shown that a system can be more reliable than its weakest (least reliable) component. Although this concept may be contrary to the old adage that a "chain is only as strong as its weakest link," it is a vital concept in understanding the advantages and disadvantages of the various configuration possibilities in system design. For example, a parallel configuration yields a higher system reliability than a series configuration if the same components are used in each design.

In the final analysis, then, system reliability can be said to be related to the following three basic areas:

1. System design configuration

2. Component reliability

3. Rate of failure

Methodologies for determining system reliability based on the three areas are presented in the sections which follow.

Series Configuration. For the constant hazard model (exponential) system reliability is found by:

$$R(t) = \prod_{i=1}^{n} e^{-\lambda_i t} = \exp -\left[\sum_{i=1}^{n} \lambda_i\right] t$$

where n = number of components in series

λ_i = failure rate of component i

In using this formulation, the following assumptions are made:

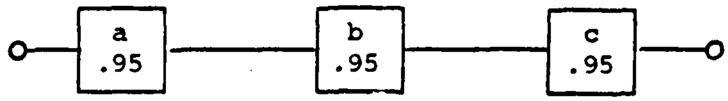
1. Components are connected in series

2. Components are independent, i.e., failure of one has no effect on another component.

EXAMPLE 20-4

An electronic device has three components with equal reliability as illustrated below.

Figure 20-18



For $t = 10$ and $\lambda_i = .0054$, find the system reliability and mean time to failure.

$$R_s = e^{-[.0054 + .0054 + .0054](10)}$$

$$= e^{-(.0162)(10)}$$

$$= 0.85^{*}$$

Note that the system reliability can also be found by using the probability law of intersection, namely that

$$R_s = R_a \cap R_b \cap R_c$$

$$\text{or} \quad R_s = R_a * R_b * R_c$$

$$= (.95)^3$$

$$= 0.85^{*}$$

*Third decimal place will differ due to rounding of λ and R_i.
MTTF is found by

$$MTTF = \frac{1}{\sum \lambda_i}$$

$$= \frac{1}{.0054 + .0054 + .0054}$$

$$= 61.7 \text{ hours}$$

For a linear increasing hazard model, system reliability can be found by:

$$= \prod_{i=1}^{n} e^{-\frac{\alpha_i t^2}{2}}$$

$$= \exp\left(-\sum_{i=1}^{n} \frac{\alpha_i t^2}{2} \right)$$

System reliability for this design is:

$$R_s(10) = 1 - \left[\prod_{i=1}^{3}\left(1 - e^{-.0054(10)}\right)\right]$$

$$= \quad 1 - (.0257)^3$$

$$= \quad 0.9998^*$$

Using the probability law for union, the same reliability is found by:

$$R_s(t) = R_a \cup R_b \cup R_c$$

$$R_s(t) = R_a + R_b + R_c - R_a \cdot R_b - R_b \cdot R_c - R_a \cdot R_c + R_a \cdot R_b \cdot R_c$$

$$= 3(.95) - 3(.95)^2 + (.95)^3$$

$$= 0.9998^*$$

* Fifth decimal place will be different due to rounding of λ and R_i.

Also using the law of union, the MTTF is found by:

$$\text{MTTF} = \frac{1}{\lambda_a} + \frac{1}{\lambda_b} + \frac{1}{\lambda_c} - \frac{1}{\lambda_a + \lambda_b} - \frac{1}{\lambda_b + \lambda_c} - \frac{1}{\lambda_a + \lambda_c} + \frac{1}{\lambda_a + \lambda_b + \lambda_c}$$

$$= 3\left(\frac{1}{.0054}\right) - 3\left(\frac{1}{.0054 + .0054}\right) + \left(\frac{1}{3(.0054)}\right)$$

$$= \text{339.51 hours}$$

System reliability can be found for the Weibull model in a similar manner. If p components have a constant hazard model and n - p components have a linear increasing hazard, system reliability can be readily found by:

$$R_s(t) = \left(\prod_{i=1}^{p} e^{-\lambda_i t}\right)\left(\prod_{i=p+1}^{n} e^{\frac{\alpha_i t^2}{2}}\right)$$

<u>Parallel Configuration</u>. For components connected in parallel, reliability for the system is determined in the following manner.

Recall from probability theory that

$$P\ (\text{failure}) = 1 - P\ (\text{success})$$

and $$R(t) = 1 - F(t)$$

For the constant hazard model, then, system reliability is:

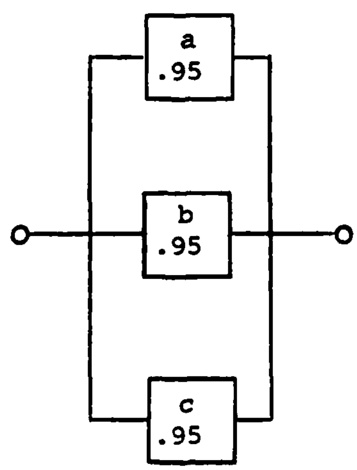
$$R_s(t) = 1 - \left[\prod_{i=1}^{n}\left(1 - e^{-\lambda_i t}\right)\right]$$

The mean time to failure is found by using the union law for probability.

<u>EXAMPLE 20-5</u>

Using the same three components as in the example for the system in series, the following parallel design results:

Figure 20-19



MTTF can also be found by:

$$\text{MTTF}^{*} = \frac{1}{\lambda} \cdot \left[\sum_{i=1}^{n} \frac{1}{i} \right]$$

$$= \frac{1}{.0054} \cdot \left[\frac{1}{1} + \frac{1}{2} + \frac{1}{3} \right]$$

$$= 339.51 \text{ hours}$$

*Only true for parallel components and when component reliability is equal for all components.

From the example above, it can readily be seen that both MTTF and system reliability have increased considerably over that which was achieved in the series configuration. This is due to redundancy which exists in the parallel design. The system designer must weigh the alternatives, i.e., the additional cost and weight added to the system with redundant components against the increased reliability and MTTF.

In summary, this chapter has only included a brief introduction to the basic concepts of reliability. Other topics such as reliability allocation, switching, and reliability growth are contained in publications devoted to a more complete treatment of the subject. The user should consult these prior to attempting to solve problems which are other than the most basic.

CHAPTER 21

ECONOMIC CONCEPTS

21-1 PRESENT VALUE CONCEPTS

A dollar in hand today is worth more than a claim to a dollar tomorrow--that is the essence of the time value of money. This statement implies that money has the power to earn money. Its earning power is related to the productiveness of capital investment it makes possible and the impatience on the part of those who would rather borrow to spend today than to save and spend tomorrow. Earning power is expressed as a rate of interest. Thus, money deposited in a bank savings account is said to earn interest at some stipulated rate per year or other time period.

In illustration, a bank deposit, P, earns interest at some rate, i, per interest period. At the end of one interest period, it will grow to some amount, A, according to the following formula:

$$(1) \quad P\,(1+i) = A$$

For any number of interest periods, n, P will grow to A according to the following formula:

$$(2) \quad P\,(1+I)^n = A,$$

the term in the parenthesis being referred to as the compound amount factor, or simply, CAF. We can express (2) alternatively as:

$$(3) \quad P = A\,\frac{1}{(1+i)^n},$$

where P is shown as the present value of some future amount, A, and the term $\frac{1}{(1+i)^n}$, is referred to as the present value factor, or simply, PVF. Thus, if an individual wished to accumulate some amount A at the end of n interest periods, given that the interest rate were i, it would be a simple matter to calculate the amount, P, necessary to achieve that goal. That is, P is the present value of the future amount, A.

Such a lump sum is in someways similar to a direct investment in productive capital from which flow goods or services with a value that exceeds the amount of the investment. The difference between the value of output and the investment cost of the capital is

the total return on the capital. Ordinarily, a capital asset is expected to produce output reasonably steadily over its lifetime generating a flow of revenue sufficient to provide a complete recapture of its invested cost and a return thereon. Thus, the revenue that is realized periodically from produced output is partially a return of capital and partially a return on capital. Given an interest rate or rate of return, i, that is desired for some investment, P, over its life, n, one can determine the periodic and uniform amount, R, necessary to satisfy those conditions. We can express P as the present value of the future stream of revenue according to the following formula:

$$(4) \quad P = R_1 \frac{1}{(1+i)} + R_2 \frac{1}{(1+i)^2} + R_3 \frac{1}{(1+i)^3} \cdot \cdot \cdot + R_n \frac{1}{(1+i)^n}$$

Since R is uniform, (4) can be expressed as:

$$(5) \quad P = R \left[\frac{(1+i)^n - 1}{i(1+i)^n} \right]$$

the term in brackets being referred to as the series present value factor, or simply, SPVF. Alternatively, we can express R as the amount of revenue recurring periodically for some investment, P, over its life, n, with a rate of return, i.

$$(6) \quad R = P \left[\frac{i(1+i)^n}{(1+i)^n - 1} \right]$$

The term in brackets being referred to as the capital recovery factor, or simply, CRF.

If from (4) we can calculate the present value of a future revenue stream, we can also use the formula to calculate the present value of a stream of periodic costs, C. Thus,

$$(7) \quad P = C_o + C_1 \frac{1}{(1+i)} + C_2 \frac{1}{(1+i)^2} + \cdot \cdot \cdot + C_n \frac{1}{(1+i)^n}$$

Accordingly, for a nonrevenue producing public program, we are enabled to calculate the present value of all costs by (7), providing we know the interest rate or rate of return, i, to use. For a given rate of return, we can calculate the present value of costs for each of a number of competing alternatives designed to satisfy fully stated program objectives.

The suggestion of applying the concept of the time value of money to government programs may seem farfetched indeed, and yet it has the very same significance to money spent for government purposes as it does to private enterprise. Whatever money is spent for government purposes has alternative private uses. Those alternative uses for money would produce an investment return or would satisfy someone's impatience to consume now. It is reasonable to expect government purposes to achieve results as beneficial as the alternative uses to which the money might otherwise be put.

However, a problem arises in applying the time value concept to evaluate government program alternatives in that the opportunity rate of return is open to interpretation. For those of us in DOD, however, this has already been decided. DODI 7041.3, 18 October 1972, directs that an annual interest rate of 10% shall be used to evaluate program proposals involving time periods exceeding one year. That interest rate is thought to reflect the economic character of alternative uses for the money DOD proposes to use for its purposes. Thus, a planner can use (7) to evaluate feasible alternatives that will satisfy the objectives of a DOD program over its anticipated life and find which is the least costly in present value terms.

Evaluating alternatives, however, is not as straightforward a matter as (7) might suggest. The life of an alternative proposal is only infrequently identical to that of the program it is designed to satisfy or to that of another alternative with which it is competing. In this respect it is important to understand that competing alternatives must be evaluated over the same time interval. Thus, program life is the correct time interval for evaluating each alternative proposal. This means that for alternatives with lives shorter than program life, the present value of replacing an original investment in hardware and facilities together with the present value of any residual items at the end of the program must be taken into consideration. For an alternative with a life greater than the program life, the present value of residuals at the end of the program are recognized in like manner.

Naturally, every planning problem involving the concept of present value is not related to a new program. Many such problems are related to existing programs in which the alternatives are to replace or continue use of existing facilities. Conditions under which a defense program is implemented may very well change during its life, and the means to satisfy its continuing objectives may also change. Thus, it is entirely possible that replacement of existing facilities and hardware may be economically feasible; present value analysis will confirm whether or not replacement should take place.

These two basic problems of choosing among competing alternatives to satisfy the objectives of a new program, and choosing between an in-place alternative and a proposed replacement for an already implemented program are illustrated in the following examples.

EXAMPLE 21-1:

A U. S. Navy shipyard is to be equipped with new metal working facilities to satisfy increased fleet requirements anticipated over the next 30 years. Three alternatives, each with a different useful life are being considered. Their investment costs and annual operating and maintenance costs are shown below. None of the alternatives will have a salvage value at the end of its life.

Table 21-1

	Alternatives 1	2	3
Life	5	10	15
Required Investment	\$1,000	\$2,700	\$3,500
Annual O&M Costs	600	500	200

Present value of program life cycle costs at 10% interest factor

Investment + PVF of Annual Costs over Program Life

Alternative 1: $\$1,000 + 600\ (SPVF_5) + PVF_5\ [1000 + 600\ (SPVF_5)]$

$+ PVF_{10}\ [1000 + 600\ (SPVF_5)] + \ldots + PVF_{25}\ [1000 + 600$

$(SPVF_5] = 1000 + 600\ (3.791) + 1.487\ (3275) = \$8135$

Alternative 2: $\$2,700 + 500\ (SPVF_{10}) + (PVF_{10} + PVF_{20})\ [2700 + 500$

$(SPVF_{10})]$

$2700 + 500\ (6.145) + (.535)[2700 + 500\ (6.145)] = \$8865$

Alternative 3: $\$3,500 + 200\ (SPVF_{15}) + (PVF_{15})\ [3500 + 200\ (SPVF_{15})]$

$3500 + 200\ (7.606 + .239\ [3500 + 200\ (7.606)] = \$6215$

We observe that the first alternative, if chosen, would be replaced five times during the 30-year program life, and that the second alternative would be replaced twice, while the third alternative would be replaced once. In each case, the present value of replacement investments has been calculated. We have assumed zero salvage value for each of the alternatives, but if there were salvage value, its present value should be calculated appropriately and subtracted from the present value of all alternative costs.

From the above calculations, we can see that Alternative 3 with its higher investment cost but lower O&M costs has the lowest present value life cycle costs.

EXAMPLE 21-2:

The transient aircraft repair facility at a certain U.S. Air Force Base has become inadequate to accommodate recent vintage jet aircraft. Because of the inadequacy, long delays are commonplace in servicing and repair operations. It has been decided to improve the facility either by modification or replacement. The existing facility has a salvage value of approximately $90,000. Modification investment costs are estimated to be $100,000. Annual operating costs of the modified facility are estimated to be $60,000. Alternatively, the facility could be replaced with a new plant costing $300,000 but for which annual operating costs would be only $35,000. The rejuvenated facility would have an expected life of 20 years, after which it would have no salvage value. The replacement facility is estimated to have a salvage value of $80,000 at the end of 20 years.

Present value of program life cycle costs at 10% interest factor

Modified Facility: $\$190{,}000 + 60{,}000\ (SPVF_{20})$

$= 190{,}000 + 60{,}000\ (8.5136) \qquad = \$700{,}816$

Replacement Facility: $\$300{,}000 + 35{,}000\ (SPVF_{20})$

$- 80{,}000\ (PVF_{20})$

$= 300{,}000 + 35{,}000\ (8.5136)$

$- 80{,}000\ (.1486) \qquad = \$586{,}088$

We observe that modifying the existing facility would be the more expensive alternative even though it would require a lesser investment to make it the operational equal of the replacement alternative.

21.2 MINIMIZING VARIABLE COSTS TO PRODUCE A GIVEN LEVEL OF OUTPUT IN TWO ALTERNATIVE PRODUCTION FACILITIES

In producing any single item at some given level of output, the manager of defense production facilities seeks to do so at minimum total cost. The available resources with which to produce output may very well include two or more fixed facilities each capable by itself of meeting the production goal. Should the manager accordingly choose one or another facility based on which has the least average cost to produce the given level of output?

Somewhat surprisingly, minimum cost to produce can usually be achieved by allocating output among all the alternatives simultaneously. Least cost production requires that output be allocated among alternative facilities so that their marginal costs are equal. Marginal cost, MC, of course, is the rate of change in total cost, TC, with respect to the rate of change in output, x. Thus,

$$MC = \frac{d(TC)}{dx}$$

Accordingly, the invested cost of an alternative facility has nothing whatsoever to do with the problem. It is there and already paid for; the only question is how variable resources such as labor and material that are subject to control can be combined with the fixed facilities to produce output in the least costly manner. Thus, a manager of in-place facilities, in evaluating their use, need concern himself with variable costs only.

The concept of minimum cost output can be illustrated in the following terms. Two production facilities, A and B, each have the capacity to produce a given level (say 10) of output of a single item. The unique total variable cost (TVC) curve, average variable cost (AVC) curve and marginal cost (MC) curve of each is displayed in Figure 21-1.

It can be seen that average cost to produce in A falls initially to some minimum value then rises, but for any level of output greater than 10 units is less than average cost in B for the same output. This would seemingly suggest that to produce any output greater than 10 units at minimum cost, facility A should be used.

But a glance at the marginal cost curves of the two facilities tells us that the marginal cost at the given level of output in A exceeds marginal cost for lesser levels of output in B. If some of a total number to be produced were reallocated to B so that both facilities produce simultaneously, would total cost be reduced? Yes, indeed. And so long as MC in B were less than MC in A, total cost could be reduced further by successive reallocation to B from A. When MC in B were equal to A, minimum cost allocation would be achieved.

The reason that total cost is reduced when MC in B is less than in A is that by a reallocation of one unit from A to B, total cost is reduced more in A than it is increased in B. In more general terms, the minimum cost allocation for any output level, Q, would be as shown in Figure 21-2 where $MC_A = MC_B$.

Figure 21-1

TOTAL VARIABLE COST

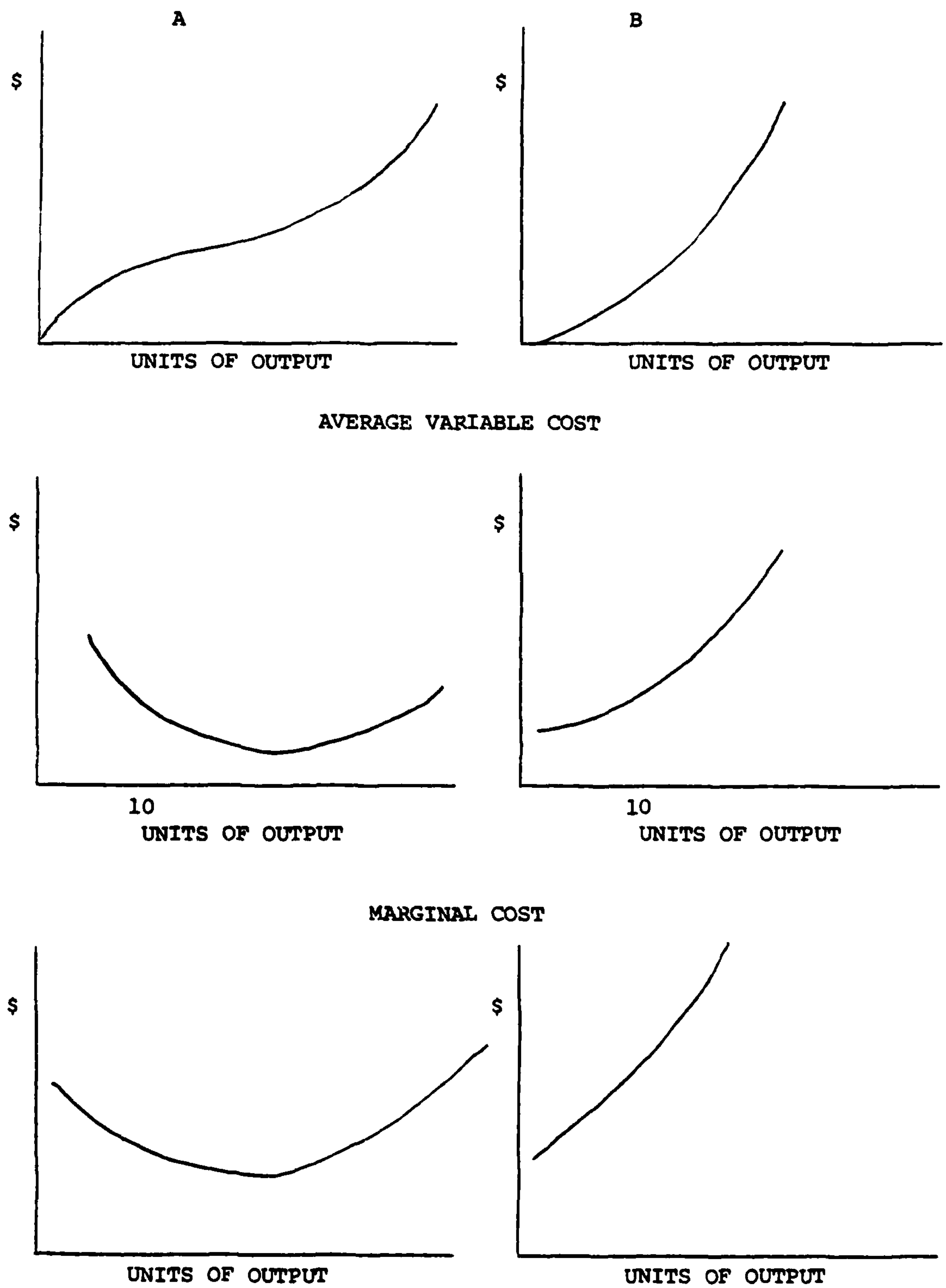
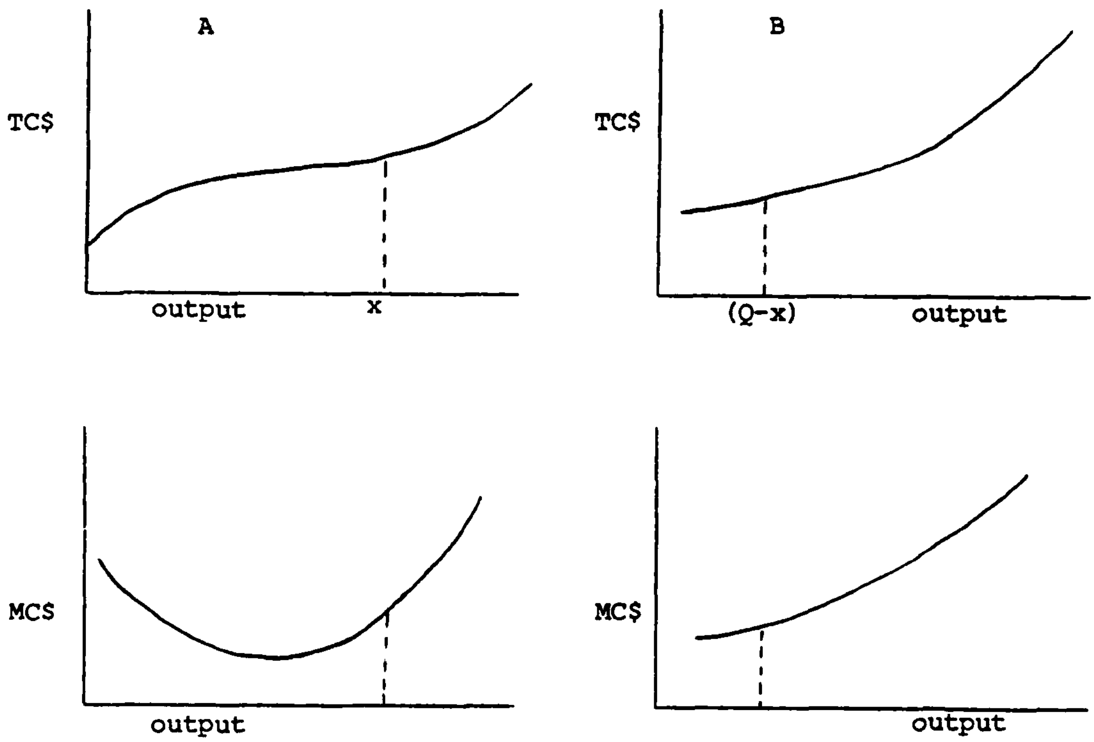


Figure 21-2



Naturally, the total cost to produce the required level of output is the sum of the total cost in A to produce x units and the total cost in B to produce (Q-x) units.

The cost and output characteristics of production facilities can sometimes be fitted to algebraic curves as total variable cost functions. Where this is possible, the solution for minimum cost allocation can be determined by mathematical techniques. To illustrate the procedure, let us assume that the two production facilities A and B have the following total variable cost functions

$$TVC_A = \frac{1}{2}x^3 - 50x^2 + 4610x$$

$$TVC_B = 3x^3 + 180x^2 + 2000x$$

where x represents the unit of output.

A level of 70 units of output must be produced. Each facility, remember, has the capacity to produce the 70 units by itself. Output is to be allocated between them so as to minimize cost. In order to solve for minimum cost output, we must first combine the two functions into one common expression. Thus,

$$TVC = TVC_A + TVC_B$$

But in doing this, we must remember that of a total output of 70, the number allocated to A is not allocated to B. Accordingly, we take x as the variable of output in A and substitute for the variable of output in B, the variable, (70 - x). The total variable cost function now becomes

$$TVC = \frac{1}{2}x^3 - 50x^2 - 4610x + 3(70-x)^3 + 180(70-x)^2 + 2000(70-x)$$

We know that to minimize cost

$$MC = \frac{d(TVC)}{dx} = 0 \text{ and } \frac{d(MC)}{dx} > 0$$

In differentiating TVC, collecting terms and setting the result equal to zero, we have

$$MC = 3x^2 - 608x + 26676 = 0$$

Solving this quadratic equation, we get

$$x \approx 64, 138$$

Testing the second derivative at the two solution values of x, we have

$$\frac{d(MC)}{dx} \text{ for } 64 > 0; \quad \frac{d(MC)}{dx} \text{ for } 138 < 0$$

From calculus we know that a relative minimum value of the dependent variable (in this case, TVC) exists if the second derivative is positive when evaluated at some value of the independent variable (in this case, x

Accordingly, our allocation for minimum cost output must be

Facility A: ≈ 64

Facility B: $(70-x) \approx 6$

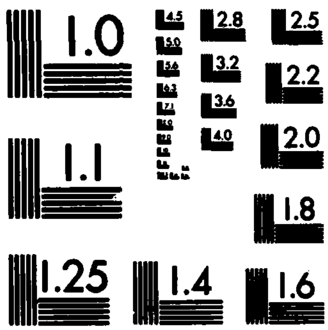
AD-A121 990   QUANTITATIVE TOOLS FOR THE LOGISTICS MANAGER(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS J E ENGEL APR 80 AFIT-LS-32   7/7

UNCLASSIFIED   F/G 15/5   NL

END

FILMED

DTIC

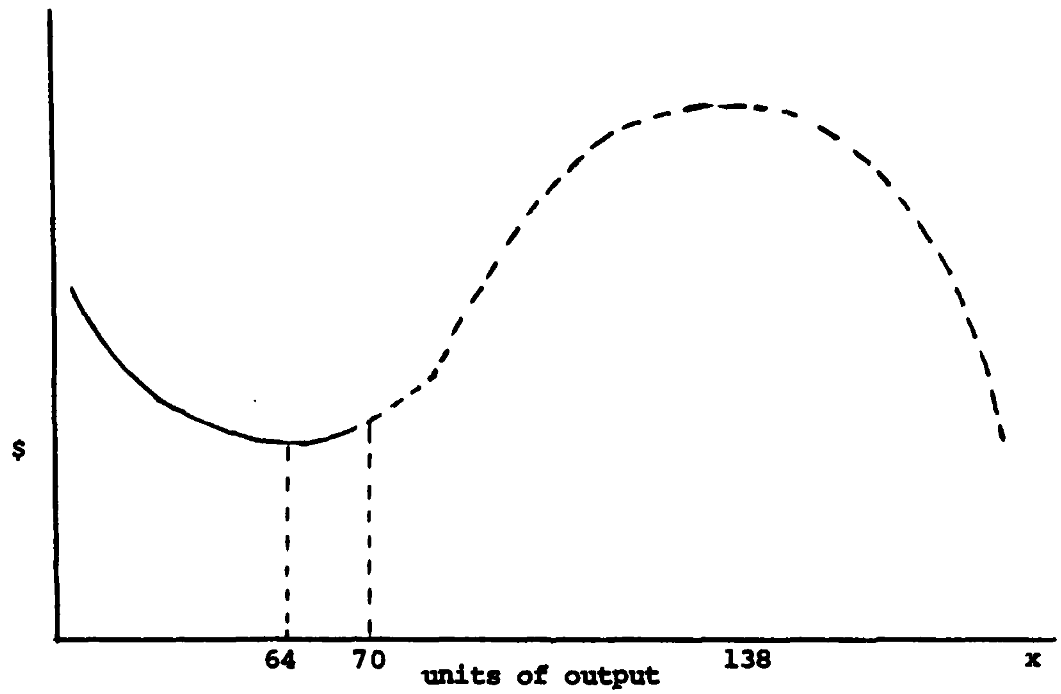


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

The question arises why a total variable cost curve which should be expected always to rise for increasing output should, in this example, have an unseemly relative minimum value and a relative maximum value. The curve, $TVC = TVC_A + TVC_B$ when plotted in terms of the variable of production in A looks like

Figure 21-3

TOTAL VARIABLE COST
(in terms of the variable, x, in facility A)



We can observe that it falls from an initially high value to a relative minimum at approximately 64 units of output, according to our solution, as more and more relatively high MC units in B are reallocated as relatively low MC units in A. But further allocation from B to A in excess of 64 results in increasing total variable cost as marginal cost in A for each reallocated unit exceeds the marginal cost in B. Beyond 70 units of output, the curve has no relevance to our problem, since for that dictated level, any allocation greater than 70 to A would require a fictitious allocation of negative units to B.

Two considerations suggest that a further development of least cost allocation can be made. First, it is possible to make a plot of the allocation between two production facilities against required levels of total output, making it easy to determine allocation for whatever level of output is imposed on the facilities. Alternatively, a

computer program can be developed for the allocation. Second, if the number of alternative facilities exceeds two, minimum cost allocation still requires equality of marginal costs; however, the solution process is more complicated and is the subject for a later treatment of this basic idea.

Often, the production process in a facility does not lend itself to analytical expression, and so least cost allocation is a trial and error proposition. Thus, in making an initial allocation of variable labor and materials among alternative facilities, the production manager may rely primarily on a combination of intuition and experience.

In refining the initial allocation, he should seek to reduce total cost for each successive allocation. So long as the cost to produce an additional unit in one facility is less than that of other facilities, output in that facility should be increased by a reduction in output in others; total variable cost will fall. When MC is the same in each facility, total variable cost can be reduced no further.

CHAPTER 22

NONPARAMETRIC STATISTICS

INTRODUCTION

Frequently it is desired to conduct statistical analyses under conditions which do not permit the assumption of a parametric technique to be made or the requirement of such a technique to be met. In most instances, there exists a corresponding nonparametric or "distribution free" statistic which, although generally less powerful, may be employed since the assumptions are less restrictive. In this chapter a representative selection of these nonparametric statistics, chosen to include those which are the most useful to the logistician, is presented. They are grouped into three major classifications: measures and tests involving nominal (categorical) level data, measures and tests involving ordinal level data, and goodness of fit tests.

MEASURES AND TESTS INVOLVING NOMINAL (CATEGORICAL) LEVEL DATA

A. A One-Sample Test of Occurrence Frequencies/Probabilities: Chi Square Row Vector.

General description/purpose: A chi square row vector may be used to test whether a statistically significant difference exists between an observed number of subjects, objects, or responses falling into each of a set of categories (cells) and an expected number based on hypothesized frequencies/probabilities of occurrence within the categories

Assumptions/limitations:

1. The null hypothesis (H_0) completely states the proportions of objects falling into the categories (cells) of the presumed population so that the expected frequencies can be deduced.

2. Category (cell) size limitations (criteria):

a. If the degrees of freedom (df) equals 1, the expected frequency of each category (cell) (E_i; $i = 1,2,\ldots,c$) should be ≥ 5.

If df ≥ 2, each E_i should be ≥ 1 and at least 80% of the E_is should be ≥ 5.

b. Expected frequencies may be increased by combining adjacent catagories (cells), but this is desirable only if such combinations can be meaningfully made.

c. If either

(1) One starts with only two categories (cells) and has one (or both) $E_i < 5$,

(2) or, if after combining adjacent categories (cells), one ends up with only two categories (cells) and still has one (or both) $E_i < 5$, then the binomial test (see "Other test(s)") should be used rather than the chi square.

Discussion/methodology: The procedure for application of the chi square row vector test involves these steps:

1. Cast the observed frequency of each category (cell) (0_i; $i = 1,2, \ldots, c$) into its corresponding component of a c-dimensional row vector. The sum of the frequencies should be N, the number of independent observations.

2. Determine the degrees of freedom: df = c − 1.

3. From H_o, determine the expected frequency of each category (cell):

$$E = \begin{cases} \text{hypothesized frequency of category (cell) } i, \\ \text{as stated in } H_o \\ \qquad\qquad \text{or} \\ (\text{hypothesized probability of category (cell) } i, \\ \text{as stated in } H_o) \cdot N \end{cases}$$

$$i = 1, 2, \ldots, c$$

When df ≥ 2, if more than 20% of the E_i's are < 5, combine adjacent categories (cells), where this is reasonable, until the category (cell) size limitations (criteria) are met. (Note: The value of c will be reduced, while the values of some of the E_i's will be increased.) When df = 1, use this test (approximately) only if both E_is are ≥ 5.

The data may be conveniently displayed as follows:

Categories				Total
0_1 / E_1	0_2 / E_2	...	0_c / E_c	N

where $N = \sum_{i=1}^{c} 0_i$

4. Compute the value of the test statistic, χ^2 observed or χ_o^2:

$$\chi_o^2 = \begin{cases} \sum_{i=1}^{c=2} \frac{(|O_i - E_i| - \frac{1}{2})^2}{E_i} & \text{if df} = 1 \ (c=2) \\ \text{or} & \\ \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i} & \text{if df} \geq 2 \ (c \geq 3) \end{cases}$$

5. Determine the critical value for the test statistic, χ^2 critical or χ_o^2:

$$\chi_c^2 = \begin{cases} \chi^2_{\alpha,df} = \chi^2_{\alpha},\ c-1 & \text{if a two-tailed test at the } \alpha\text{-level of significance} \\ \chi^2_{2\alpha,df} = \chi_{2\alpha,c-1} & \text{if a one-tailed test at the } \alpha\text{-level of significance} \end{cases}$$

6. Resolve the hypothesis test: if $\chi_o^2 > \chi_c^2$, reject H_o.

EXAMPLE 22-1

The AF-wide mix of enlisted personnel in a particular logistics career field is 83% male and 17% female. At Boon Docks AFB, 79 enlisted men and 21 enlisted women in this career field are assigned. Is the mix of enlisted personnel at this base typical of the career field? Test at $\alpha = 0.10$.

H_o: The male (M)/female (F) mix of enlisted personnel at Boon Docks AFB is typical of the particular logistics career field, i.e., $P(M) = 0.83$ and $P(F) = 0.17$.

H_1: It is atypical, i.e., $P(M) \neq 0.83$ and $P(F) \neq 0.17$

Male(M)	Female(F)	N
79 / 83	21 / 17	100

$$df = c-1 = 2-1 = 1$$

$$E_M = P(M)*N = (0.83)(100) = 83$$

$$E_F = P(F)*N = (0.17)(100) = 17$$

E_M and E_F both ≥ 5? (yes)

$$\chi_o^2 = \sum_{i=1}^{c=2} \frac{(|O_i - E_i| - \frac{1}{2})^2}{E_i} \qquad \text{since df} = 1$$

$$= \frac{(|79-83|-\frac{1}{2})^2}{83} + \frac{(|21-17|-\frac{1}{2})^2}{17}$$

$$\cong 0.87$$

$$\chi_c^2 = \chi_{\alpha,df}^2 = \chi_{.10,\,1}^2 = 2.71$$

$$\chi_o^2 \cong 0.87 < 2.71 = \chi_c^2$$

Conclusion: Fail to reject H_o. There is insufficient evidence to conclude that the male/female mix of enlisted personnel at this base is atypical of the particular logistics career field at the $\alpha = 0.10$ level of significance.

Reference(s). Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. (McGraw-Hill: New York, 1956) pp. 42-47.

Other test(s). Binomial test, see Siegel, pp. 36-42.

B. The Two Related Samples Case: The McNemar Test for Difference or Changes

General description/purpose: The McNemar test may be used to test whether a statistically significant difference exists between two treatments, or whether one treatment is "better" than another, when each subject "serves as his own control" by being exposed to both treatments at different times. It is particularly applicable to "before and after" treatment designs in which measurement is in the strength, of either a nominal or an ordinal scale to assess the "before to after" change, in direction only, due to a single treatment.

Assumptions/limitations:

1. The usual parametric technique for analyzing data from two related samples is the application of the T-test to the difference scores. There are, however, circumstances under which the paired-samples t-test is inapplicable:

a. The assumptions/requirements of the paired-samples t-test are considered unrealistic or are found not to hold for the data.

b. It is preferable to avoid making the assumptions or testing the requirements of the paired-samples t-test and, thus, give greater generality to the conclusions drawn (at the expense of employing a less powerful test).

c. The differences between matched pairs are not represented as scores, but rather as "signs", i.e., it is indicated which member of any pair is "greater than" the other, but the magnitude of the difference is not known.

d. The score data are simply classificatory, i.e., the two members of any matched pair can either respond in the same way or in entirely different ways which do not stand in any order or quantitative relation.

In these instances, the McNemar test may be applied

2. If the expected frequency of changes [½(A + D)] is < 5, the binomial test (see reference under "Other test(s)") should be used rather than the McNemar test.

Discussion/methodology:

The procedure for application of the McNemar test involves these steps:

1. Cast the observed frequencies into a fourfold table of the form:

		After +	After -
Before	-	A	B
	+	C	D

where A is the number of subjects who changed their responses from - to + as a result of the application of the treatment, etc.

2. Determine the expected frequency of change: $E = \frac{1}{2}(A + D)$. If $E < 5$ use the binomial test rather than the McNemar test.

3. Compute the value of the test statistic χ^2 observed or χ^2_o:

$$\chi^2_o = \frac{(|A-D|-1)^2}{A + D}$$

4. Determine the critical value for the test statistic, χ^2 critical or χ^2_o:

$$
\chi_c^2 = \begin{cases} \chi^2_{\alpha,1} & \text{if a two-tailed test at the } \alpha\text{-level of significance} \\ & \text{or} \\ \chi^2_{2\alpha,1} & \text{if a one-tailed test at the } \alpha\text{-level of significance} \end{cases}
$$

5. Resolve the hypothesis test: if $\chi_o^2 > \chi_c^2$, reject H_o.

EXAMPLE 22-2

Currently, single enlisted personnel in overseas areas residing in dormitories and subsisting in government dining facilities are not authorized a cost of living allowance (COLA), which is paid to other co-assigned military personnel. A proposal to fund singles' COLA was first submitted to the Congress in DOD's FY79 supplemental budget request. To explain the rationale for requesting singles' COLA, HQ USAF/DCS Manpower and Personnel produced a Palace Flick on the subject. As part of a test of the potential of this Palace Flick to influence the attitudes of targeted groups of key military personnel concerning singles' COLA, a sample of 25 junior majors without overseas unit command experience and assigned to the Pentagon was selected. Each subject was questioned about his attitude concerning singles' COLA, shown the Palace Flick, and then questioned again about his attitude. The attitude data from the sample were as follows:

		Attitude concerning singles' COLA after being shown the Palace Flick	
		Favored	No opinion or did not favor
Attitude concerning singles' COLA before being shown the Palace Flick	No opinion or did not favor	14	3
	Favored	6	2

Is the Palace Flick effective in favorably (positively) influencing (changing) the attitudes of junior majors without overseas unit command experience concerning singles' COLA? Test at $\alpha = 0.01$.

H_o: For those junior majors without overseas unit command experience whose attitudes concerning singles' COLA are influenced by being shown the Palace Flick, the probability that any major will change his attitude from "No opinion or does not favor" to "Favor" (P_A) is equal to the probability that he will change his attitude from "Favors" to "No opinion or does not favor" (P_D) is equal to ½. i.e. $P_A = P_D = ½$,

H_1: $P_A > P_D$

$E = \frac{1}{2}(A + D) = \frac{1}{2}(14 + 2) = 8$ $\qquad$ $E \geq 5$? Yes

$$\chi_o^2 = \frac{(|A-D|-1)^2}{A + D}$$

$$= \frac{(|14-2|-1)^2}{14 + 2}$$

$$\doteq 7.56$$

Since this is a one-tailed test at $\alpha = 0.01$ with rejection region determined by H_1: $P_A > P_D$, it corresponds to a two-tailed test at $2\alpha = 0.02$ with rejection region determined by H_1: $P_A \neq P_D$.

$$\chi_c^2 = \chi_{2\alpha,1}^2 = \chi_{0.02,\ 1}^2 = 5.41$$

$$\chi_o^2 \doteq 7.56 > 5.41 = \chi_c^2$$

Conclusion: Reject H_o. Conclude that the Palace Flick is effective in favorably (positively) influencing (changing) the attitudes of junior majors without overseas unit command experience concerning singles' COLA at the $\alpha = 0.01$ level of significance.

Reference(s). Siegel, pp. 63-67.

Other test(s). Binomial test; see Siegel, pp. 36-42.

C. The Two Independent Samples Case. A test of Occurence Frequencies/ Probabilities or for Independence: <u>Chi Square Contingency Table</u>.

General description/purpose: A chi square contingency table may be used to test whether a statistically significant difference exists between two independent groups with respect to some characteristic; specifically, between observed numbers of subjects, objects, or responses falling into each of a set of categories (cells) from each of two groups and the expected frequencies of occurrence within the categories based on the hypothesis of independence between the groups with respect to the characteristic.

Assumptions/limitations:

1. If the observed frequencies of occurrence consistitute a 2 x 2 contingency table (two categories (cells) for each of the two groups being contrasted), then the possible application of the chi square contingency table test should be determined on the basis of the following criteria:

a. When the number of observations $N > 40$, use this test with the test statistic corrected for continuity (see step 4 in "Discussion/methodology" below).

b. When $20 \leq N \leq 40$, use this test with continuity correction <u>only if</u> the expected frequency of <u>each</u> category (cell) (E_{ij}; $i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, c$) is ≥ 5. If the smallest expected frequency is < 5, use the Fisher exact probability test (see reference under "Other test(s)").

c. When $N < 20$, always use the Fisher test.

2. If the degrees of freedom (df) is ≥ 2, use this test <u>only if</u> each $E_{ij} \geq 1$ <u>and</u> at least 80% of the E_{ij}'s are ≥ 5 (category (cell) size limitations (criteria).

Expected frequencies may be increased by combining adjacent categories (cells), but this is desirable only if such combinations can be meaningfully made.

3. When df ≥ 2, this test is insensitive to the effects of order. If, then, hypotheses take order into account, this test may not be the best. For methods that strengthen the chi square tests when H_0 is tested against specific (including ordered) alternatives, consult Cochran (1954) (see "References").

Discussion/methodology:

The procedure for application of the chi square contingency table test involves these steps:

1. Cast the observed frequency of each category (cell) (O_{ij}, $i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, c$) into its corresponding position in an $(r \times c)$ - dimensional array. The assignments of the groups and the values of the characteristics to either the rows or the columns of the array is immaterial. The sum of the frequencies should be N, the number of independent observations.

2. Determine the degrees of freedom; $df = (r-1)(c-1)$.

3. Determine the expected frequency of each category (cell) as the product of its marginal totals divided by N:

$$E_{ij} = \frac{(\text{marginal total of row } i)*(\text{marginal total of column } j)}{N}$$

$$i = 1, 2, \ldots, r; \quad j = 1, 2, \ldots, c$$

When df ≥ 2, if more than 20% of the E_{ij}'s are <5, combine adjacent categories (cells), where this is reasonable, until the category (cell) size limitations (criteria) are met. (Note: The values of r and/or c will be reduced, while the values of some of the E_{ij}'s will be increased.)

The data may be conveniently displayed as follows:

Groups or Values of the Characteristic				Marginal Totals	
O_{11} E_{11}	O_{12} E_{12}	. . . .	O_{1c} E_{1c}	$\sum_{j=1}^{c} O_{1j}$	
$\vdots$					
O_{r1} E_{r1}	O_{r2} E_{r2}	. . . .	O_{rc} E_{rc}	$\sum_{j=1}^{c} O_{rj}$	
$\sum_{i=1}^{r} O_{i1}$	$\sum_{i=1}^{r} O_{i2}$		$\sum_{i=1}^{r} O_{ic}$	N	Total

Where $E_{ij} = \dfrac{\left(\sum_{k=1}^{c} O_{ik}\right) * \left(\sum_{k=1}^{r} O_{kj}\right)}{N}$ and $N = \sum_{i=1}^{r} \sum_{j=1}^{c} O_{ij}$

4. Compute the value of the test statistics, χ^2 observed or χ^2_o:

$$\chi^2_o = \begin{cases} \displaystyle\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(|O_{ij}-E_{ij}|-\frac{1}{2})^2}{E_{ij}} & \text{if df} = 1 \ (r = 2 \text{ and } c = 2) \\ \text{or} & \\ \displaystyle\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \left[\sum_{i=1}^{r} \sum_{j=1}^{c} \left(\frac{O_{ij}}{E_{ij}} \right) \right] - N & \text{if df} \geq 2 \end{cases}$$

5. Determine the critical value for the test statistic, χ^2 critical or χ^2_o:

$$\chi_o^2 = \begin{cases} \chi^2_{\alpha,df} = \chi^2_{\alpha,\ (r-1)(c-1)} & \text{if a two-tailed test at the } \alpha\text{-level of significance} \\ \text{or} & \\ \chi^2_{2\alpha,df} = \chi^2_{2\alpha,(r-1)(c-1)} & \text{if a one-tailed test at the } \alpha\text{-level of significance} \end{cases}$$

6. Resolve the hypothesis test: if $\chi_o^2 > \chi_c^2$, reject H_o.

EXAMPLE 22-3

To increase safety-consciousness and reduce the incidence of work-related accidents throughout the command, AFLC is considering whether all personnel should be tested annually on safety procedures and regulations. In preparation for this testing, it has been proposed that all personnel annually attend a safety refresher mini-course. To investigate the effect of attendance at such a mini-course on personnel performance on a safety test, the headquarters staff developed three (3) safety refresher mini-courses with two (2), four (4), and eight (8) hours of instruction and drafted a test on safety procedures and regulations. At an unidentified ALC, 200 randomly selected persons attended either one or none of the mini-courses and then were administered the safety test. The test results were as follows:

		Number of hours of instruction attended in safety refresher mini-course			
		No instruction	2 hours	4 hours	8 hours
Performance on safety test	PASS	9	50	38	18
	FAIL	21	50	12	2

On the basis of the outcomes of this experiment, should AFLC personnel's performance on a test on safety procedures and regulations be expected to be affected by the number of hours of instruction attended in a safety refresher mini-course? Test at $\alpha = 0.05$.

H_o: AFLC personnel's performance on a test on safety procedures and regulations is independent of the number of hours of instruction attended in a safety refresher mini-course.

H_1: It is dependent thereon.

Number of hours of instruction attended in safety refresher mini-course

Performance on Safety Test		No instruction	2 hours	4 hours	8 hours	Marginal Totals
	PASS	$\boxed{9}$ 17.25	$\boxed{50}$ 57.50	$\boxed{38}$ 28.75	$\boxed{18}$ 11.50	115
	FAIL	$\boxed{21}$ 12.75	$\boxed{50}$ 42.50	$\boxed{12}$ 21.25	$\boxed{2}$ 8.50	85
MARGINAL TOTALS		30	100	50	20	200

$r = 2,\ c = 4 \qquad df = (r-1)(c-1) = (2-1)(4-1) = 3$

$$E_{11} = \frac{\text{(Marginal total of row 1)*(Marginal total of column 1)}}{N}$$

$$= \frac{(115)(30)}{200} = 17.25\text{, etc.}$$

Since df $\geq$ 2, is each $E_{ij} \geq 1$ and are at least 80% of the E_{ij}s ≥ 5? Yes

$$\chi_o^2 = \sum_{i=1}^{r=2} \sum_{j=1}^{c=4} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$= \underbrace{\frac{(9-17.5)^2}{7.5}}_{(1,1)\text{ cell}} + \underbrace{\frac{(50-57.5)^2}{57.5}}_{(1,2)\text{ cell}} + \ldots + \underbrace{\frac{(2-8.5)^2}{8.5}}_{(2,4)\text{ cell}}$$

$$= \left[\sum_{i=1}^{r=2} \sum_{j=1}^{c=4} \left(\frac{O_{ij}^2}{E_{ij}} \right) \right] - N$$

$$= \left[\underbrace{\frac{(9)^2}{17.5}}_{(1,1)\text{ cell}} + \underbrace{\frac{(50)^2}{57.5}}_{(1,2)\text{ cell}} + \ldots + \underbrace{\frac{(2)^2}{8.5}}_{(2,4)\text{ cell}} \right] - 200$$

$\cong 27.23$

$$\chi_c^2 = \chi_{\alpha,df}^2 = \chi_{0.05,3}^2 = 7.81$$

$$\chi_o^2 \cong 27.23 > 7.81 = \chi_c^2$$

Conclusion: Reject H_o. Strongly conclude that AFLC personnel's performance on a test on safety procedures and regulations should be expected to be affected by (dependent on) the number of hours of instruction attended in a safety refresher mini-course at the $\alpha = 0.05$ level of significance.

Reference(s). Siegel, pp. 104-111

Other test(s). Fisher exact probability test; see Siegel, pp. 96-104

Measures of Correlation: Pearson's Contingency Coefficient

General description/purpose: Pearson's contingency coefficient C is a measure of the extent of association or relation between two independent groups with respect to some characteristic, i.e., between two sets of attributes. It is uniquely useful when there is only nominal (categorical) level information available concerning one or both groups, such as when that information consists of unordered series of occurrence frequencies. The use of this contingency coefficient, in contrast to Pearson's parametric product moment (sample correlation coefficient r), does not require either (1) the assumptions of underlying continuity or a specific population distribution for the categories which comprise either or both sets of attributes, or (2) that the categories be orderable in any particular way. The latter insures that this contingency coefficient, as computed from a contingency table, will have the same value, regardless of the arrangements of the categories (cells) in the rows and columns of the table. The coefficient's value always lies between 0 and 1, and equals 0 in the case of complete independence between the groups with respect to the characteristic, i.e., when $\chi_o^2 = 0$.

Assumptions/limitations:

Pearson's C suffers from four limitations:

1. The data must be amenable to computation of the chi square test statistic (χ^2 observed or χ_o^2) before C may appropriately be used.

2. Even though C = 0 in the case of complete independence, i.e., when $\chi_o^2 = 0$, C cannot attain its upper limit of 1; in the case of complete association, the value of C depends on the numbers of rows and columns in the contingency table.

3. Two nonzero Pearson's contingency coefficients are not comparable unless they are yielded by contingency tables of the same size.

4. C is not directly comparable to any other measure of correlation, such as Pearson's r, Spearman's r_s, or Kendall's τ.

Discussion/methodology:

1. Despite the above limitations, C is an extremely useful measure of association because of wide applicability resulting from its freedom from assumptions and other restrictive requirements. C may be used to indicate the degree of relation between two sets of data when other, especially parametric, measures of association are inapplicable.

2. Pearson's contingency coefficient C is defined as follows:

$$C = \sqrt{\frac{\chi^2 \text{ observed}}{N + \chi^2 \text{ observed}}}$$

EXAMPLE 22-4

The previous example of a chi square contingency table is continued.

$$C = \sqrt{\frac{\chi^2 \text{ observed}}{N + \chi^2 \text{ observed}}}$$

$$\cong \sqrt{\frac{27.23}{200 + 27.23}}$$

$$\cong 0.3462$$

The correlation expressed by Pearson's contingency coefficient, between AFLC personnel's performance on a test on safety procedures and regulations and the number of hours of instruction attended in a safety refresher mini-course is $C \cong 0.35$.

Reference(s).

1. Everitt, B. S. The Analysis of Contingency Tables. (Halstead Press: New York, 1977) pp. 56-57.

2. Siegel, pp. 196-202.

Cramer's Contingency Coefficient

General description/purpose: Like Pearson's contingency coefficient C, Cramer's C is also a measure of the extent of association or relation between two independent groups with respect to some characteristic, i.e., between two sets of attributes. Generally, this contingency coefficient may be similarly described as Pearson's C. In particular, its value also lies between 0 and 1, and equals 0 in the case of complete independence between the groups with respect to the characteristic, i.e., when $\chi_o^2 = 0$.

Assumptions/limitations:

Cramer's C exhibits three properties similar to the limitations on Pearson's C:

1. The data must be amenable to computation of the chi square test statistic (χ^2 observed or χ_o^2) before Cramer's C may appropriately be used.

2. Cramer's C, however, in contrast to Pearson's C, may attain its upper limit of 1 in the case of complete association for all numbers of rows and columns in the contingency table.

3. Cramer's C is also not directly comparable to any other measure of correlation, such as Pearson's r, Spearman's r_s, Kendalls τ, or even Pearson's C.

Discussion/methodology:

Cramer's contingency coefficient is defined as follows:

$$\text{Cramer's C} = \frac{\chi^2 \text{ observed}/N}{\min\ (r-1,\ c-1)}$$

EXAMPLE 22-5

The previous example of a chi square contingency table is continued.

$$\text{Cramer's C} = \frac{\chi^2 \text{ observed}/N}{\min\ (r-1,\ c-1)}$$

$$\cong \frac{27.23/200}{\min\ (2-1,\ 4-1)}$$

$$\cong 0.1362$$

The correlation, expressed by Cramer's contingency coefficient, between AFLC personnel's performance on a test on safety procedures and regulations and the number of hours of instruction attended in a safety refresher mini-course is $\cong$ 0.14. (Note that this correlation, expressed by Pearson's contingency coefficient, is C $\cong$ 0.35; but, these values are not directly comparable.)

Reference(s): Everitt, pp. 56-57.

MEASURES AND TESTS INVOLVING ORDINAL LEVEL DATA

A. Related Samples Tests For Differences or Changes:

A two-sample Test: Wilcoxon Matched-Pairs Signed-Ranks Test

General description/purpose: The Wilcoxon matched pairs signed-ranks test may be used to test whether a statistically significant mean difference exists between the response levels of two treatments, or whether one treatment is "better" than another, when each subject "serves as his own control" by being exposed to both treatments at different times. It is particularly applicable to "before and after" treatment designs in which measurement is in the strength of either an interval or a ratio scale to assess the "before to after" change, in both (relative) magnitude and direction, due to a single treatment.

Assumptions/limitations

1. The usual parametric technique for analyzing data from two related samples for differences or changes is the application of the t-test to the difference scores. There are, however, circumstances under which the paired-samples t-test is inapplicable:

a. The assumptions/requirements of the paired-samples t-test are considered unrealistic or are found not to hold for the data.

b. It is preferable to avoid making the assumptions or testing the requirements of the paired-samples t-test and, thus, give greater generality to the conclusions drawn (at the expense of employing a less powerful test).

2. For n matched pairs of observations, (X_i, Y_i):

a. The measurement scale for the Xs and the Ys is either interval or ratio.

b. The pairs, (X_i, Y_i), $i = 1, \ldots, n$ constitute a random sample from a bivariate population distribution.

3. For the difference, D_i, between X_i and Y_i for the ith pair of observations:

a. Each D_i is a continuous random variable.

b. The distribution of each D_i is symmetric.

Discussion/methodology:

The procedure for application of the Wilcoxon matched-pairs signed-ranks test involves these steps:

1. Determine the differences, $D_i = X_i - Y_i$, $i = 1, \ldots, n$.

2. If any $D_i = 0$, drop it from the set and decrease n by one (1).

3. Rank order the absolute values of the differences, $|D_i|$. If ties occur, average the ranks of the absolute differences involved in the tie and assign the average as the rank of each tied value.

4. Suffix to each rank the algebraic sign of the difference, D_i, corresponding to it.

5. Determine the total of the positive signed-ranks, R+, and the total of the negative signed-ranks, R-.

6. Determine the value of the test statistic, R: R = smaller of R+, R-.

7. Determine from a table the critical value for the test statistic, $W_{(critical)}$ or W_c, and the appropriate algebraic sign of the smaller sum of ranks:

$$W_c = \begin{cases} W_{\alpha/2} & \text{(Lower tail critical value) if a two-tailed test at the } \alpha\text{-level of significance (AND R = either R- or R+)} \\ W_{\alpha} & \text{(Lower tail critical value) if a one-tailed test at the } \alpha\text{-level of significance AND R = R+ if a one-tailed test to the left, R = R- if a one-tailed test to the right} \end{cases}$$

8. Resolve the hypothesis test: if $R < W_c$, and the smaller sum of ranks is of the appropriate algebraic sign for a one-tailed test, reject H_o.

<u>EXAMPLE 22-6</u>

Ten USAF enlistees were randomly selected while undergoing Basic Military Training (BMT) to determine whether a recently-devised speed-reading program can improve their reading rates. The enlistees were given standardized reading exams before and after the program. The exam scores in words per minute were as follows:

Enlistee	Before	After
1	200	225
2	150	375
3	300	650
4	400	380
5	120	100
6	250	250
7	320	410
8	175	180
9	100	130
10	500	600

Is the recently-devised speed-reading program effective in improving (increasing) the reading rates of enlistees during BMT? Test at $\alpha = 0.05$.

$H_0 = \mu$ before $\geq \mu$ after

$H_1 = \mu$ before $< \mu$ after

The differences, D_i, the ranks of $|D_i|$, and the algebraic signs of these ranks are:

Enlistee	D_i		Rank of $\lvert D_i \rvert$	Sign
1	-25		4	-
2	-225		8	-
3	-350		9	-
4	+20	tie	2.5	+
5	+20			+
6	0		(eliminate)	
7	-90		6	-
8	-5		1	-
9	-30		5	-
10	-100		7	-

Since $D_6 = 0$, eliminate it. Now $n = 10 - 1 = 9$

$R+ = 2.5 + 2.5 = 5$

$R- = 4 + 8 + 9 + 6 + 1 + 5 + 7 = 40$

R = smaller of $R+$, $R- = R+ = 5$

Since this is a one-tailed test to the left at $\alpha = 0.05$, it is required that

$R \stackrel{?}{=} R+$. Yes

$W_c = W_{0.05,\, n=9} = 9$ from Table 22-1

$R = R+ = 5 < 9 = W_c$

Conclusion: Reject H_0. Conclude that the recently-devised speed-reading program is effective in improving (increasing) the reading rates of enlistees during BMT at the $\alpha = 0.05$ level of significance.

Reference(s). Siegel, pp. 75-83.

Other test(s). For large samples, especially for n beyond the range of tabulated critical values, there is a normal approximation; see Siegel, pp. 79-83.

Table 22-1

Critical Values of the Wilcoxon Matched-Pairs Signed-Ranks Test Statistic

Sample size n	$W_{.005}$	$W_{.01}$	$W_{.025}$	$W_{.05}$	$W_{.10}$	$W_{.20}$	n(n+1)/2
4	0	0	0	0	1	3	10
5	0	0	0	1	3	4	15
6	0	0	1	3	4	6	21
7	0	1	3	4	6	9	28
8	1	2	4	6	9	12	36
9	2	4	6	9	11	15	45
10	4	6	9	11	15	19	55
11	6	8	11	14	18	23	66
12	8	10	14	18	22	28	78
13	10	13	18	22	27	33	91
14	13	16	22	26	32	39	105
15	16	20	26	31	37	45	120
16	20	24	30	36	43	51	136
17	24	28	35	42	49	58	153
18	28	33	41	48	56	66	171
19	33	38	47	54	63	74	190
20	38	44	53	61	70	82	210

Table entries are lower tail critical values

Source: Adapted from R. L. McCormack, "Extended Tables of the Wilcoxon Matched Pairs Signed Rank Statistics," Journal of the American Statistical Association, 60 (1965), pp. 864-71.

A k-Sample Test: Friedman 2-way ANOVA by Ranks Test

General Description/purpose: The Friedman 2-way analysis of variance (ANOVA) by ranks test may be used to test whether statistically significant differences exist among k populations from which matched samples, each having the same number of cases, have been drawn. The matching may be achieved by studying the same group of n subjects under each of k conditions, or n matched groups of k subjects may be obtained, and then one subject in each group randomly assigned to the first condition, one subject in each group to the first condition, one subject in each group to the second condition, etc., through the kth condition. Observations are cast into a 2-way table having n rows, representing the various individual subjects or matched groups of subjects, and k columns, representing the various conditions. If the scores of subjects serving under all conditions are under study, then each row of the table presents the scores of one individual subject or one matched group of subjects under each of the k conditions. The data for this test are ranks. The scores for each subject, i.e., in each row, are ranked separately. So, when k conditions are being studied, the ranks for each subject range from 1 to k. The Friedman test determines

whether it is likely that the k columns of ranks (k matched samples) were drawn from the same population.

Assumptions/limitations:

1. The usual parametric technique for analyzing data from k related samples is the application of the 2-way ANOVA with k treatments (conditions) and n cases (subjects). There are, however, circumstances under which the 2-way ANOVA is inapplicable:

a. The assumptions/requirements of the 2-way ANOVA are considered unrealistic or are found not to hold for the data.

b. It is preferable to avoid making the assumptions or testing the requirements of the 2-way ANOVA and, thus, give greater generality to the conclusions drawn (at the expense of employing a less powerful test).

2. The observations from the k matched samples are in at least an ordinal scale.

Discussion/methodology:

The procedures for application of the Friedman 2-way ANOVA by ranks test involves these steps:

1. Cast the observations (scores) into a 2-way table having k columns (conditions) and n rows (individual subjects or matched groups of subjects).

2. Rank the observations (scores) in each row from 1 to k. It is immaterial whether the ranking is from lowest to highest or from highest to lowest; however, the ranking must be consistently applied across all n rows.

3. Determine the sum of the ranks in each column, R_j.

4. Determine the value of the test statistic, X^2 ranks or X_r^2, when no tied ranks are permitted:

$$X_r^2 = \left[\frac{12}{nk(k+1)} \sum_{j=1}^{k} (R_j)^2 \right] -3n(k+1)$$

where n = number of rows (subjects)

k = number of columns (conditions)

R_j = sum of ranks in jth column

$\sum_{j=1}^{k}$ directs the summation of the squares of the sums of the ranks over all k conditions

5. Determine the critical value for the test statistic, $\chi^2_{critical}$ or χ^2_c:

$$\chi^2_c = \chi^2_{\alpha,\ k-1}$$

6. Resolve the hypothesis test: if $\chi^2_r > \chi^2_c$, reject H_o.

<u>EXAMPLE 22-7</u>

Four chief master sergeants are finalists in the competition for AFLC Senior Superintendent of the Year. After reviewing their military service records and personally interviewing them, a panel of five officers, serving as judges for the competition, ranked the finalists. The ranks assigned to the finalists by the officer-judges were as follows:

		Finalists			
		CMSgt A	CMSgt B	CMSgt C	CMSgt D
	1	3	1	2	4
	2	1	2	4	3
Officer-Judges	3	3	1	2	4
	4	3	2	1	4
	5	4	1	2	3

Do the five officer-judges agree in their rankings of the four finalists? Test at $\alpha = 0.10$. If the officer-judges agree, what is their overall (composite) rank order of the four finalists?

H_o = There is no agreement among the five officer-judges in their rankings of the four finalists.

H_1 = There is agreement.

The sums of the ranks in the columns, R_j, are:

$$R_j = \sum_{i=1}^{n} R_{ij} \qquad n = 5;\ j = 1, \ldots, k;\ k = 4$$

$R_1 = 3 + 1 + 1 + 3 + 3 + 4 = 14$ (CMSgt A)

$R_2 = 7$ (CMSgt B)

$R_3 = 11$ (CMSgt C)

$R_4 = 18$ (CMSgt D)

$$\chi_r^2 = \left[\frac{12}{r\, k\, (k+1)} \sum_{j=1}^{k} (R_j)^2\right] - 3n(k+1)$$

$$= \left\{\frac{12}{(5)(4)(5)} \left[(14)^2 (7)^2 + (11)^2 + (18)^2\right]\right\} - (3)(5)(5)$$

$$= (0.12)(690) - 75 = 82.8 - 75$$

$$= 7.8$$

$$\chi_c^2 = \chi_{\alpha,k-1}^2 = \chi_{0.10,3}^2 = 6.25$$

$$\chi_r^2 = 7.8 > 6.25 = \chi_c^2$$

Conclusion: Reject H_0. Conclude that there is agreement among the five officer-judges in their rankings of the four finalists, i.e., the five ranks assigned by the officer-judges to each of the four finalists were drawn from different populations, at the $\alpha = 0.10$ level of significance.

The average ranks, $\bar{R}_j$, where

$$\bar{R}_j = \frac{R_j}{n} = \frac{1}{n} \sum_{i=1}^{n} R_{ij}$$

or, equivalently, the sums of the ranks, R_j, may now themselves be arranged in ascending order to evidence the panel's overall (composite) rank order.

Overall (composite) Rank	Finalist	$\bar{R}_j$	R_j	
1st	CMSgt B	1.4	7	Lowest
2nd	CMSgt C	2.2	11	
3rd	CMSgt A	2.8	14	
4th	CMSgt D	3.6	18	Highest

Reference(s)

1. Siegel, pp. 166-172.

2. Winer, B. J., Statistical Principles in Experimental Design, McGraw-Hill, New York, 1962, pp. 136-137.

Other Test(s)

1. For certain cases in which n and k are not large, the exact probabilities(prob-values) of occurrence are tabulated for values of the test statistic χ^2 ranks as large as an observed (computed) χ_r^2, see Siegel.

2. For the determination of an appropriate test statistic when tied ranks are permitted, see Winer, pp. 136-137.

3. For Cochran Q test for k related samples, an extension of the McNemar test for two related samples, provides a method to test whether three or more matched sets of frequencies or proportions (samples) are statistically significantly different among themselves. The matching may again be based on either the fact that the same individual subjects are employed under different conditions or on relevant characteristics of different subjects by which matched groups may be defined. The Cochran test is particularly suitable when the observations are in a nominal (categorical) scale or constitute dichotomized (two-valued) ordinal information, e.g., a manager's preference between two alternative decisions and/or courses of action. See Siegel, pp. 161-166.

B. Independent Samples Tests of Central Tendency

A Two-Sample Test: (Wilcoxon)-Mann-Whitney Test.

General description/purpose: The (Wilcoxon)-Mann-Whitney test is designed to assist in determining whether the distributions of two independent populations are the same or different. If it is assumed that any difference between the two population distributions is due only to a difference in location of the two distributions, then the (Wilcoxon)-Mann-Whitney test may be used to test whether a statistically significant difference exists between the central tendencies of the two distributions, as reflected by their medians.

Assumptions/limitations:

1. The usual parametric technique for analyzing data from two independent samples, from populations whose variances are assumed equal, for a difference in central tendencies is the application of the t-test to the difference of their means. There are, however, circumstances under which the two independent samples t-test is inapplicable:

a. The assumptions/requirements of the two independent samples t-test are considered unrealistic or are found not to hold for the data.

b. It is preferable to avoid making the assumptions or testing the requirements of the two independent samples t-test and, thus, give greater generality to the conclusions drawn (at the expense of employing a less powerful test).

2. For n_A observations, X_{A_i}, from population A and n_B observations, X_{B_i}, from population B:

a. The measurement scale for the X_As and the X_Bs is at least ordinal.

b. X_{A_i}, $i = 1, \ldots, n_A$ and X_{B_i}, $i = 1, \ldots, n_B$ constitute independent random samples.

c. The sample random variables, X_A and X_B, are continuous.

Discussion/methodology. The procedure for application of the (Wilcoxon)-Mann-Whitney test involves these steps:

1. Arrange in ascending order the $n_A + n_B$ observations from the combined samples.

2. Assign the ranks $1, \ldots, n_A + n_B$ to the ordered observations. If ties occur, average the ranks of the observations involved in the tie and assign the average as the rank of each tied value.

3. Determine the sum of the ranks assigned to the X_{A_i}, S_A, and the sum of the ranks assigned to the X_{B_i}, S_B:

$$S_A = \sum_{i=1}^{n_A} R(X_{A_i})$$

where $R(X_{A_i})$ is the rank assigned to the observation X_{A_i}, and

$$S_B = \sum_{i=1}^{n_B} R(X_{B_i})$$

similarly.

(Note: As a computational check: $S_A + S_B = \dfrac{(n_A + n_B)\ (n_A + n_B + 1)}{2}$

4. Determine the value of the test statistic, T:

a. If a two-tailed test T = either T_A or T_B, where

$$T_A = S_A - \left[\frac{n_A(n_A+1)}{2}\right]$$

$$T_B = S_B - \left[\frac{n_B(n_B+1)}{2}\right]$$

b. If a one-tailed test, T = whichever of T_A or T_B (as defined above) corresponds to the population with the smaller median as hypothesized under H_1, i.e., (1) if H_1: median (X_A) < median (X_B), $T = T_A$. (2) If H_1: median (X_A) > median (X_B), $T = T_B$.

5. Determine the critical value(s) for the test statistic, $W_{critical}$ of W_C:

a. If a two-tailed test at the α-level of significance,

$$W_C = \begin{cases} W_{\alpha/2} & \text{(lower tail critical value from a table, and} \\ W_{1-\alpha/2} & \text{(upper tail critical value)} \\ = n_A n_B - W_{\alpha/2} & \end{cases}$$

b. If a one-tailed test at the α-level of significance, $W_C = W_\alpha$ (lower tail critical value) from a table.

b. Resolve the hypothesis test:

a. If a two-tailed test, if either $T < W_{\alpha/2}$ or $T > W_{1-\alpha/2}$, reject H_o.

b. If a one-tailed test, if $T < W_\alpha$, reject H_o.

<u>EXAMPLE 22-8</u>

A chief of maintenance desires to know whether a difference exists between the times to failure of a certain electronics component from two competing manufacturers. From field tests, times to failure, in hours, of the component from the two manufacturers, A and B, were obtained as follows:

A: 76, 104, 125, 73, 87, 69, 149, 78, 101

B: 97, 88, 92, 101, 83, 71, 90, 94, 91, 85, 111

Do the average times to failure of the component differ between the two manufacturers? Test at $\alpha = 0.05$.

X_A = random variable, time to failure in hours of the component from manufacturer A.

X_B = similarly defined

H_o: median (X_A) = median (X_B)

H_1: median $(X_A) \neq$ median (X_B)

The rank-ordered, ascending arrangement of the combined samples, with ties denoted by an overbar (-), is:

Rank:	1	2	3	4	5	6	7	8	9	10
Manufacturer:	A	B	A	A	A	B	B	A	B	B
Observation:	69	71	73	76	78	83	85	87	88	90

Rank:	11	12	13	14	15.5	15.5	17	18	19	20
Manufacturer:	B	B	B	B	A	B	A	B	A	A
Observation:	91	92	94	97	101	101	104	111	125	149

These ties assigned average of tied ranks.

$$S_A = \sum_{i=1}^{n_A} R(X_{A_i}) \qquad n_A = 9$$

$$= 1 + 3 + 4 + \ldots + 20 = 92.5$$

$$S_B = \sum_{i=1}^{n_B} R(X_{B_i}) \qquad n_B = 11$$

$$= 2 + 6 + 7 + \ldots + 18 = 117.5$$

Check: $S_A + S_B = 92.5 + 117.5 = 210 \stackrel{?}{=}$

$$\frac{(n_A + n_B)(n_A + n_B + 1)}{2} = \frac{(20)(21)}{2} = 210 \qquad \text{yes}$$

$$T_A = S_A - \frac{n_A(n_A + 1)}{2}$$

$$= 92.5 - \frac{(9)(10)}{2} = 92.5 - 45$$

$$= 47.5$$

$$T_B = S_B - \frac{n_B(n_B + 1)}{2}$$

$$= 117.5 - \frac{(11)(12)}{2} = 117.5 - 66$$

$$= 51.5$$

Since this is a two-tailed test, T = either T_A or T_B.

$$W_C = \begin{cases} W_{\alpha/2} = W_{0.025;\ n_A = 9,\ n_B = 11} = 24 \\ \quad \text{from Table (K), page 276} \\ \quad \text{of Siegel, and} \\ W_{1-\alpha/2} = W_{0.975;\ n_A = 9,\ n_B = 11} \end{cases}$$

$$= n_A n_B - W_{\alpha/2}$$

$$= (9)(11) - 23 = 99 - 24 = 75$$

If t = either T_A or T_B:

$$W_{\alpha/2} = 24 < 47.5 = T_A,\ 51.5 = T_B < 76 = W_{1-\alpha/2}$$

i.e., neither $T < W_{\alpha/2}$ nor $T > W_{1-\alpha/2}$

Conclusion: Fail to reject H_o. There is insufficient sample evidence to conclude that the average times to failure in hours of the component differ between the two manufacturers at the $\alpha = 0.05$ level of significance.

Reference(s). Siegel, pp. 116-127.

Other test(s):

1. For large samples, expecially for n_As and n_Bs beyond the ranges of tabulated critical values, there is a normal approximation, with a correction for excessive numbers of ties; see Siegel, pp. 121-126.

2. The median test also provides a method to test whether two independent populations differ in central tendencies and may also be employed whenever observations are in at least an ordinal scale of measurement. However, the median test is almost always less powerful than the (Wilcoxon)-Mann-Whitney test. Additionally, it possesses the unattractive feature that its power-efficiency (the percent of time the test will concur in the conclusion of rejection of the null hypothesis with the parametric t-test, when both are applied to data which meet the requirements for analysis with the parametric t-test) decreases as the sample sizes increase. See Siegel, pp. 111-116.

A k-Sample Test: <u>Kruskal-Wallis 1-Way ANOVA by Ranks Test</u>

General description/purpose: The Kruskal-Wallis 1-way analysis of variance (ANOVA) by ranks test is designed to assist in determining whether the distributions of k independent populations are the same or different.

If it is assumed that any differences among the k population distributions are due only to differences in location of the k distributions, then the Kruskal-Wallis test may be used to test whether statistically significant differences exist among the central tendencies of the k distributions, as reflected by their medians.

Assumptions/limitations:

1. The usual parametric technique for analyzing data from k independent samples, from populations whose variances are assumed equal, for differences in central tendencies is the application of the single factor, completely randomized design analysis of variance (1-way ANOVA). There are, however, circumstances under which the 1-way ANOVA is inapplicable:

a. The assumptions/requirements of the 1-way ANOVA are considered unrealistic or are found not to hold for the data.

b. It is preferable to avoid making the assumptions or testing the requirements of the 1-way ANOVA and, thus, give greater generality to the conclusions drawn (at the expense of employing a less powerful test).

2. For n_1 observations, X_{1_i}, from population 1, n_2 observations, X_{2_i}, from population 2, ..., n_k, observations X_{k_i}, from population k.

a. The measurement scale for the X_j, $j = 1, \ldots, k$ is at least ordinal.

b. X_{1_i}, $i = 1, \ldots, n_1$, X_{2_i}, $i = 1, \ldots, n_2$; ..., X_{k_i}, $i = 1, \ldots, n_k$ constitute mutually independent random samples.

c. The sample random variables, X_j, $j = 1, \ldots, k$, are continuous.

Discussion/methodology:

The procedure for application of the Kruskal-Wallis test involves these steps:

1. Arrange in ascending order the $n_1 + n_2 + \ldots + n_k$ observations from the combined samples.

2. Assign the ranks $1, \ldots, n_1 + n_2 + \ldots + n_k$ to the ordered observations. If ties occur average the ranks of the observations involved in the tie and assign the average as the rank of each tied value.

3. Determine the sum of the ranks assigned to the X_{1_i}, R_1, the sum of the ranks assigned to the X_{2_i}, R_2, ..., and the sum of the ranks assigned to the X_{k_i}, R_k:

$$R_j = \sum_{i=1}^{n_j} R(X_{j_i}) \qquad j = 1, \ldots, k$$

where $R(X_{j_i})$ is the rank assigned to the observation X_{j_i}.

(Note. As a computational check:

$$\sum_{j=1}^{k} R_j = \frac{N(N+1)}{2}$$

where $N = n_1 + n_2 + \ldots + n_k$)

4. Determine the value of the test statistic, T:

$$T = \frac{12}{N(N+1)} \sum_{j=1}^{k} \left\{ \frac{\left[R_j - (\tfrac{1}{2})n_j \quad (N+1)\right]^2}{n_j} \right\}$$

$$= \left\{ \frac{12}{N(N+1)} \sum_{j=1}^{k} \left[\frac{(R_j)^2}{n_j} \right] \right\} - 3(N+1)$$

where $N = n_1 + n_2 + \ldots + n_k$

5. Determine the critical value for the test statistic, $T_{critical}$ or T_c:

$$T_c = \begin{cases} T_{\approx\alpha} & \text{from a table whenever } k, n_1, n_2 \ldots, n_k \text{ are within range, or} \\ \chi^2_{\alpha,k-1} & \text{(approximately) otherwise} \end{cases}$$

6. Resolve the hypothesis test: if $T > T_c$, reject H_o.

EXAMPLE 22-9

Three types of wire cord are being considered, with which to secure a certain heavy piece of aerospace ground equipment to cargo loading pallets for transport by a C-5. Samples of the three types of wire cord were subjected to a test of breaking strength. The measured forced (in coded units) required to break the wire cords were as follows:

Type 1: 250, 220, 315, 260, 205
Type 2: 185, 230, 215, 245
Type 3: 260, 285, 245, 250, 320

Do the average breaking strengths differ among the three types of wire cord? Test at $\alpha \approx 0.05$.

X_j = random variable, breaking of type j wire cord, j = 1, 2, 3

H_o: The median breaking strengths of the three types of wire cord are equal.

H_1: They are not equal, i.e., at least one type of wire cord has a greater median breaking strength than the others.

The rank-ordered, ascending arrangement of the combined samples, with ties denoted by an overbar (-), is:

Rank:	1	2	3	4	5	6.5	6.5
Type:	2	1	2	1	2	2	3
Observation:	185	205	215	220	230	$\overline{245}$	$\overline{245}$

Rank:	8.5	8.5	10.5	10.5	12	13	14
Type	1	3	1	3	3	1	3
Observation	$\overline{250}$	$\overline{250}$	$\overline{260}$	$\overline{260}$	285	315	320

Ties assigned average of tied ranks

$$R_1 = \sum_{i=1}^{n_1} R(X_{1_i}) \qquad n_1 = 5$$

$$= 2 + 4 + 8.5 + 10.5 + 13 = 38$$

$$R_2 = \sum_{i=1}^{n_1} R(X_{2_i}) \qquad n_2 = 4$$

$$= 1 + 3 + 5 + 6.5 = 15.5$$

$$R_3 = \sum_{i=1}^{n_3} R(X_{3_i}) \qquad n_3 = 5$$

$$= 6.5 + 8.5 + 10.5 + 12 + 14 = 51.5$$

Check: $N = n_1 + n_2 + n_3$

$$= 5 + 4 + 5 = 14$$

$$R_1 + R_2 + R_3 = 38 + 15.5 + 51.5 = 105 \stackrel{?}{=}$$

$$\frac{N(N+1)}{2} = \frac{(14)(15)}{2} = 105 \qquad \text{yes}$$

$$T = \frac{12}{N(N+1)} \sum_{j=1}^{k} \left\{ \frac{\left[R_j - (\tfrac{1}{2})n_j(N+1)\right]^2}{n_j} \right\}$$

$$= \frac{12}{(14)(15)} \left\{ \frac{\left[38 - (\tfrac{1}{2})(5)(15)\right]^2}{5} \right.$$

$$+ \frac{\left[15.5 - (\tfrac{1}{2})(4)(15)\right]^2}{4}$$

$$\left. + \frac{\left[51.5 - (\tfrac{1}{2})(5)(15)\right]^2}{5} \right\}$$

$$\approx (0.057)\left[(0.05) + (52.5625) + 39.2)\right]$$

$$\approx 5.233$$

$$T_c = T_{\approx\alpha} = T_{\approx 0.05};\ k = 3,\ \underbrace{n_j = 5,\ 4,\ 5}$$

$n_1 = 5$, $n_2 = 4$, $n_3 = 5$
but the order of the sample sizes n_j does not affect the critical value determination from a table.

$T_{0.050;\ k = 3,\ n_j = 5,\ 5,\ 4} = 5.6429$ from Table (O), page 283 Siegel

$$T \approx 5.233 < 5.6420 = T_c$$

Conclusion: Fail to reject H_o. There is insufficient sample evidence to conclude that the average breaking strengths differ among the three types of wire cord, i.e., that at least one type of wire cord has a greater average breaking strength than the others, at the $\alpha \approx 0.05$ level of significance.

Reference(s). Siegel, pp. 184-193.

Other test(s)

1. There is a correction for excessive numbers of ties; see Siegel, pp. 191-192.

2. The extension of the median test also provides a method to test whether k independent populations differ in central tendencies and may also be employed whenever observations are in at least an ordinal scale. However, the extension of the median test is in essence a χ^2 test for k samples which involves a 2 x k contingency table and, hence, suffers from the limitatins of contingency table cell size criteria; See Siegel, pp. 179-184.

C. Measures and Tests of Correlation.

The Two-Sample Case: <u>Spearman Rank Correlation Coefficient Test</u>

General description/purpose. The Spearman rank Correlation coefficient test may be used to test whether a statistically significant correlation (linear association) exists between the paired values from two processes. In the case of the existence of a statistically significant correlation, the Spearman rank correlation coefficient may be used as a measure of the extent of the correlation (degree of the linear association).

Assumptions/limitations:

1. The usual parametric technique for analyzing data from two related samples for correlation is the application of the t-test to the Pearson produce-moment correlation coefficient. There are, however, circumstances under which the correlation t-test is inapplicable:

a. The assumptions/requirements of the correlation t-test are considered unrealistic or are found not to hold for the data.

b. It is preferable to avoid making the assumptions or testing the requirements of the correlation t-test and, thus, give greater generality to the conclusions drawn (at the expense of employing a less powerful test).

2. For n pairs of observations, (X_i, Y_i):

a. The measurement scale for the Xs and the Ys is at least ordinal.

b. The pairs, (X_i, Y_i), $i = 1, \ldots, n$ constitute a random sample from a bivariate population distribution of continuous random variables X and Y.

Discussion/methodology:

The procedure for application of the Spearman rank correlation coefficient test involves these steps:

1. Assign the ranks 1, ..., n in ascending order to the observations of each sample separately. If ties occur, average the ranks of the observations involved in the tie and assign the average as the rank of each tied value.

2. For each pair of observations, (X_i, Y_i), determine the difference between the ranks of the component observations, d_i:

$$d_i = R(X_i) - R(Y_i)$$

where $R(X_i)$ is the rank assigned to the observation X_i, and $R(Y_i)$ similarly.

3. Determine the value of the Spearman rank correlation coefficient, r_s, which is the test statistic:

$$r_s = 1 - \left[\frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^2\right]$$

4. Determine from a table the critical value for the test statistic, $r_{critical}$ or r_c:

$$r_c = \begin{cases} r_{\alpha/2} & \text{if a two-tailed test at the } \alpha\text{-level of significance} \\ r_\alpha & \text{if a one-tailed test at the } \alpha\text{-level of significance} \end{cases}$$

5. Resolve the hypothesis test: if $|r_s| > r_c$, reject H_0.

<u>EXAMPLE 22-10</u>

While undergoing Basic Military Training (BMT), enlistees are given both verbal skills and quantitative skills tests to assist in the selection of their career fields of assignment. To determine whether there is a correlation (linear association) between the verbal skills and the quantitative skills evidenced by enlistees tested during BMT, ten enlistees were randomly selected, but without their knowledge. Their scores on the two skills tests were as follows:

Enlistees	Verbal skills test score	Quantitative skills test score
1	90	70
2	100	60
3	75	60
4	80	80
5	60	75
6	75	90
7	85	100
8	40	75
9	95	85
10	65	65

Does there exist a correlation (linear association) between the verbal skills and the quantitative skills evidenced by enlistees tested during BMT? Test at $\alpha = 0.05$.

X = random variable, verbal skills test score of an enlistee tested during BMT

Y = random variable, quantitative skills test score...

H_o: $\rho_{XY} = 0$ (X and Y are uncorrelated)

H_1: $\rho_{XY} \neq 0$ (X and Y are correlated, i.e., either there is a tendency for larger values of x to be paired with larger values of Y (direct correlation), or there is a tendency for smaller values of X to be paired with larger values of Y (indirect or inverse correlation).

The ranked observations and the differences of these ranks, d_i, are:

Enlistee	X_i	$R(X_i)$	Y_i	$R(Y_i)$	d_i
1	90	8	70	4	4
2	100	10	60	1.5	8.5
3	75	4.5	60	1.5	3
4	80	6	80	7	-1
5	60	2	75	5.5	-3.5
6	75	4.5	90	9	-4.5
7	85	7	100	10	-3
8	40	1	75	5.5	-4.5
9	95	9	85	8	1
10	65	3	65	3	0

ties assigned average of tied ranks.

The value of the Spearman rank correlation coefficient, r_s, is:

$$r_s = 1 - \left[\frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^{\,2}\right] \qquad n = 10$$

$$= 1 - \left\{\frac{6}{(10)(99)} \left[(4)^2 + (8.5)^2 + \ldots + (0)^2\right]\right\}$$

$$= 1 - \left[\left(\frac{6}{990}\right) (161)\right] \approx 0.0242$$

$$r_c = r_{\alpha/2} = r_{0.025,\ n=10} = 0.6364 \text{ from Table 22-2}$$

$$|r_s| \approx 0.0242 < 0.6364 = r_c$$

Conclusion: Fail to reject H_o. There is insufficient sample evidence to conclude that there exists a correlation between the verbal skills and the quantitative skills evidenced by enlistees tested during BMT at the $\alpha = 0.05$ level of significance.

Reference(s). Siegel, pp. 202-213.

Other test(s).

1. There is a correction for moderate or greater numbers of ties; see Siegel, pp. 206-210.

2. For large samples, especially for n beyond the range of tabulated critical values, there is a t-test approximation; see Siegel, pp. 212.

3. The Kendall rank correlation coefficient test also provides a method to determine whether the paired values of two processes are correlated and may also be employed whenever observations are in at least an ordinal scale of measurement. The Kendall test enjoys a preference over the Spearman test for excessive numbers of ties. Since both tests utilize the same amount and kind of sample information (ranks), they are equally powerful in rejecting H_o and thereby detecting the existence of correlation (linear association). The Kendall rank correlation coefficient, τ, and the Spearman rank correlation coefficient, r_s, have different underlying scales, however. When computed from the same data, they will generally differ in numerical value; they are not directly comparable to each other. The Kendall test has an advantage over the Spearman test in that τ can be generalized to a partial correlation coefficient. See Siegel, pp. 223-229.

Table 22-2

Critical Values of the Spearman Test Statistic

n	a=.100	.050	.025	.010	.005	.001
4	.8000	.8000				
5	.7000	.8000	.9000	.9000		
6	.6000	.7714	.8286	.8857	.9429	
7	.5357	.6786	.7450	.8571	.8929	.9643
8	.5000	.6190	.7143	.8095	.8571	.9286
9	.4667	.5833	.6833	.7667	.8167	.9000
10	.4424	.5515	.6364	.7333	.7818	.8667
11	.4182	.5273	.6091	.7000	.7455	.8364
12	.3986	.4965	.5804	.6713	.7273	.8182
13	.3791	.4780	.5549	.6429	.6978	.7912
14	.3626	.4593	.5341	.6220	.6747	.7670
15	.3500	.4429	.5179	.6000	.6536	.7464
16	.3382	.4265	.5000	.5824	.6324	.7265
17	.3260	.4118	.4853	.5637	.6152	.7083
18	.3148	.3994	.4716	.5480	.5975	.6904
19	.3070	.3895	.4579	.5333	.5825	.6737
20	.2977	.3789	.4451	.5203	.5684	.6586
21	.2909	.3688	.4351	.5078	.5545	.6455
22	.2829	.3597	.4241	.4963	.5426	.6318
23	.2767	.3518	.4150	.4852	.5306	.6186
24	.2704	.3435	.4061	.4748	.5200	.6070
25	.2646	.3362	.3977	.4654	.5100	.5962
26	.2588	.3299	.3894	.4564	.5002	.5856
27	.2540	.3236	.3822	.4481	.4915	.5757
28	.2490	.3175	.3749	.4401	.4828	.5660
29	.2443	.3113	.3685	.4320	.4744	.5567
30	.2400	.3059	.3620	.4251	.4665	.5479

Source: Adapted from G. J. Glasser and R. F. Winter, "Critical Values of the Coefficient of Rank Correlation for Testing the Hypothesis of Independence," Biometrika, 48 (1961), pp. 444-48.

The k-Sample Case: <u>Kendall Coefficient of Concordance Test</u>

General description/purpose: The Kendall coefficient of concordance test may be used to test whether a statistically significant association exists among the matched values from k processes. In the case of the existence of a statistically significant association, the Kendall coefficient of concordance may be used as a measure of the degree of the association. The values of the k variables are measured in, or transformed to, ranks. When observations consist of n sets of matched values from the k variables, the Kendall coefficient of concordance test is equivalent to the Friedman 2-Way ANOVA by ranks test with the roles of k and n reversed.

Assumptions/limitations:

Same as for the Friedman test with the roles of k and n reversed. See page 22-18.

Discussion/methodology:

The procedure for application of the Kendall coefficient of concordance test involves these steps:

1. Cast the observations into a 2-way table having n columns (sets of matched values, or entities to be ranked) and k rows (processes, or judges assigning ranks).

2. Rank the observations in each row from 1 to n. It is immaterial whether the ranking is from lowest to highest or from highest to lowest, however, the ranking must be consistently applied across all k rows.

3. Determine the sum of the ranks in each column, R_j, and the mean of these rank sums, $\bar{R}_j$:

$$R_j = \frac{\sum_{j=1}^{n} R_j}{n}$$

4. Determine the value of the Kendall coefficient of concordance, W, when the proportion of ties is not large:

$$W = \frac{s}{(1/12)k^2(n^3-n)}$$

where s = sum of squares of the observed deviations from the mean of the R_j.

$$= \sum_{j=1}^{n} (R_j - \bar{R}_j)^2$$

k = number of rows (sets of rankings, or judges)

n = number of columns (entities-objects or individuals-ranked).

$(\frac{1}{12})k^2(n^3-n)$ = maximum possible sum of the squared deviations, i.e., the sum s which would occur with perfect agreement among the k sets of rankings.

Roughly speaking, W is an index of the divergence of the actual agreement evidenced by the observations from the maximum possible (perfect) agreement.

5. Determine the value of the test statistics, χ_W^2:

$$\chi_W^2 = k(n-1)W$$

When no ties are permitted, $\chi_W^2 = \chi^2_{ranks}$ of the Friedman 2-way ANOVA by ranks test.

6. Determine the critical value for the test statistic, $\chi^2_{critical}$ or χ_c^2:

$$\chi_c^2 = \chi^2_{\alpha,\ n-1}$$

7. Resolve the hypothesis test: If $\chi_W^2 > \chi_c^2$, reject H_o.

EXAMPLE 22-11

The example first resolved with the Friedman 2-way ANOVA by ranks test is not approached as a Kendall coefficient of concordance test application.

FINALISTS

		CMSgt A	CMSgt B	CMSgt C	CMSgt D
	1	3	1	2	4
	2	1	2	4	3
Officer-judges	3	3	1	2	4
	4	3	2	1	4
	5	4	1	2	3
		$R_1=14$	$R_2=7$	$R_3=11$	$R_4=18$

H_o: There is no agreement among the five officer-judges in their rankings of the four finalists.

H_1: There is agreement.

$$\bar{R}_j = \frac{\sum_{j=1}^{n} R_j}{n} \qquad n = 4$$

$$= \frac{14 + 7 + 11 + 18}{4} = 12.5$$

$$s = \sum_{j=1}^{n} (R_j - \bar{R}_j)^2$$

$$= (14 - 12.5)^2 + (7 - 12.5)^2 + (11 - 12.5)^2 + (18 - 12.5)^2$$

$$= 2.25 + 30.25 + 2.25 + 30.25 = 65$$

$$W = \frac{s}{(1/12)k^2(n^3-n)} \qquad k = 5,\ n = 4$$

$$= \frac{65}{(1/12)(5^2)[(4)^3-4]}$$

$$= \frac{65}{(1/12)(25)(60)} = 0.52$$

$$\chi_W^2 = k(n-1)W$$

$$= (5)(4-1)(0.52)$$

$$= 7.8 = \chi^2_{\text{ranks}}$$

$$\chi_c^2 = \chi^2_{\alpha,n-1} = \chi^2_{0.10,3} = 6.25$$

$$\chi_W^2 = 7.8 > 6.25 = \chi_c^2$$

Conclusion: Reject H_o, etc.

Reference(s). Siegel, pp. 229-238

Other test(s).

1. For certain cases in which n is not large, the critical values of s are tabulated for selected k and levels of significance. Table (R), page 286 of Siegel, includes the cases for n = 3 to 7, selected k between 3 and 20, and α = 0.01 and 0.05.

2. There is a correction for excessive numbers of ties; see Siegel, pp. 233-235.

GOODNESS OF FIT (GOF) TESTS

A. One-Sample Tests Against an Hypothesized Distribution.

Chi Square GOF Test

General description/purpose: The chi square GOF test may be used to test whether (a) statistically significant difference(s) exist(s) between an empirical distribution and an hypothesized theoretical distribution, either discrete or continuous, the parameter(s) of which may be prespecified or estimated from sample observations. The test is an application of a chi square row vector.

Assumptions/limitations:

1. The discussion on page 22-1 applies.

2. When the hypothesized theoretical distribution is discrete, each value of that random variable may constitute a cell. However, since the test must meet the cell size criteria of a chi square row vector, some combining of adjacent cells may be necessary.

3. When the hypothesized theoretical distribution is continuous, arbitrary cells must be created by the investigator. Each cell consists of a range of values which that random variable can take on. The choices of upper and lower limits for each cell are somewhat arbitrary and require that subjective determinations be made by the investigator.

Discussion/methodology:

The discussion on page 22-1 applies, except that:

1. df = k - 1 - NEP

Where k = number of cells

NEP = number of estimated parameters.

2. $\chi^2_{critical}$ or $\chi_c^{\ 2} = \chi^2_{\alpha,df}$.

3. The test is a one-tailed test to the right; i.e., if $\chi_o^{\ 2} > \chi_c^{\ 2}$, reject H_o.

EXAMPLE 22-12

(Discrete distribution). The chief of the base pharmacy believes that customers arrive for service in accordance with a Poisson process. To test his belief, the number of arrivals was observed for 710 randomly selected 5 minute periods, as follows:

No. of arrivals within 5 minute time period:	0	1	2	3	4	5	6	7 or more
No. of time periods observed:	50	100	150	155	110	65	60	20

Do these data support the base pharmacy chief's belief that customer's arrivals are Poisson? Test at $\alpha = 0.05$.

X = random variable, the number of arrivals within a five minute time period at the base pharmacy.

$H_o = X \sim \text{Poisson } (\lambda = \underline{3.0})$

$H_1 = X \not\sim \text{Poisson } (\lambda = \underline{3.0})$ $\lambda = \hat{\lambda}$, estimated from sample

No. of Arrivals within 5 minute time period (X)	0	1	2	3	4	5	6	7 or more	Total
No. of time periods observed (O_i)	50	100	150	155	110	65	60	20	710=n
$P(X=i)=p_i$	0.05	0.149	0.224	0.224	0.168	0.101	0.05	0.034	1.0
No. of time periods expected ($E_i=p_i*n$)	36	106	159	159	119	72	36	24	711

$$\hat{\lambda} = \overline{X} = \frac{\sum_{i=0}^{7} X_i f_i}{n} = \frac{(0)(50) + (1)(100) + (2)(150) + \ldots + (7)(20)}{710}$$

$$\cong 3.0 \text{ where } f_i = 0_i$$

$$\chi_o^2 \text{ (or } \chi_s^2) = \sum_{i=0}^{7} \left[\frac{(0_i - E_i)^2}{E_i}\right] = \frac{(50-36)^2}{36} + \frac{(100-106)^2}{106} + \frac{(150-159)^2}{159} +$$

$$\ldots + \frac{(20-24)^2}{24} \cong 24.47$$

df = (no. of cells) - 1 - (no. estimated parameters) = k - 1 - NEP

= 8 - 1 - 1 = 6

$$\chi_c^{\,2} = \chi^2_{0.05,\,6} = 12.6$$

$$\chi_o^2 \cong 24.47 > 12.6 = \chi_c^{\,2}$$

Conclusion: Reject H_o. Customer arrivals at the base pharmacy are not distributed Poisson ($^\circ\lambda$ = 3.0) at the 5% level of significance.

(Note: Since λ = 3.0 was estimated from the observations, this value is the "best" choice for the parameter λ of a candidate-Poisson. Consequently, as Poisson (λ=3.0) was rejected, no other Poisson distribution should be expected to evidence a goodness of fit.

EXAMPLE 22-13

(Continuous distribution). It is suspected that the scores of USAF enlistees on a particular aptitude test administered during Basic Military Training (BMT) are normally distributed. The scores of 100 randomly selected enlistees were as follows (blocked data):

Score on a particular aptitude test	No. of enlistees who achieved score
60 - 62	5
63 - 65	18
66 - 68	42
69 - 71	27
72 - 74	8

Do these data support the suspicion that enlistee's scores on a particular aptitude test are normally distributed? Test at $\alpha = 0.05$:

x = random variable, the score of an enlistee on a particular aptitude test (integer value).

H_0: $X \sim$ normal (μ = $\underline{67.45}$, σ^2 = $\underline{(2.92)^2}$)

H_1: $X \not\sim$ normal (μ = $\underline{67.45}$, σ^2 = $\underline{(2.92)^2}$) $\bar{X} = \hat{\mu}$, $s^2 = \hat{\sigma}^2$, estimated from sample

i	Class Interval	Midpoint M_i	Class Bound B_i	Observed O_i	Z_{B_i}	$P(Z_{B_{i-1}} \leq Z \leq Z_{B_i}) = P_i$	Expected $E_i = P_i * n$	$\frac{(O_i - E_i)^2}{E_i}$
			$-\infty$					
1	60-62	61		5		0.0446	4.46	0.0654
			62.5		-1.7			
2	63-65	64		18		0.2514 - 0.0446 = 0.2068	20.68	0.3473
			65.5		-0.67			
3	66-68	67		42		0.6406 - 0.2514 = 0.3892	38.92	0.2437
			68.5		+0.36			
4	69-71	70		27		0.9177 - 0.6406 = 0.2771	27.71	0.0182
			71.5		+1.39			
k=5	72-74	73		8		1.0 - 0.9177 = 0.0823	8.23	0.0064
			∞					
				n = 100			100	0.681

$$\hat{\mu} = \overline{X} = \frac{1}{n} \sum_{i=1}^{k} M_i f_i = \frac{1}{100} [(61)(5) + (64)(18) + (67)(42) + (70)(27) +$$

$$(73)(8)] \cong 67.45 \text{ where } f_i = O_i$$

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^{k} (M_i - \overline{X})^2 f_i = \frac{1}{(100-1)} [(61) - 67.45)^2(5) + (64 - 67.45)^2(18)$$

$$+ \ldots + (73 - 67.45)^2(8)] \cong 8.526 \cong (2.92)^2$$

$$Z_{B_i} = \frac{B_i - \overline{X}}{s}; \quad \text{so } Z_{B_i} = \frac{62.5-67.45}{2.92} \cong -1.7, \text{ etc.}$$

$$\chi^2_0 \text{ (or } \chi^2_s) = \sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i}\right] \cong 0.681$$

$$df = k - 1 - NEP = 5 - 1 - 2 = 2$$

$$\chi^2_c = \chi^2_{0.05,\ 2} = 5.99$$

$$\chi^2_o \quad 0.681 < 5.99 = \chi^2_c$$

Conclusion. Fail to reject H_0. There is insufficient sample evidence to conclude that the scores of enlistees on a particular aptitude test administered during BMT are not distributed normal ($\mu = 67.45$, $\sigma^2 = (2.92)^2$) at the 5% level of significance. (See accompanying illustration in Figure 22-1.

Reference(s). Siegel, pp. 42-47.

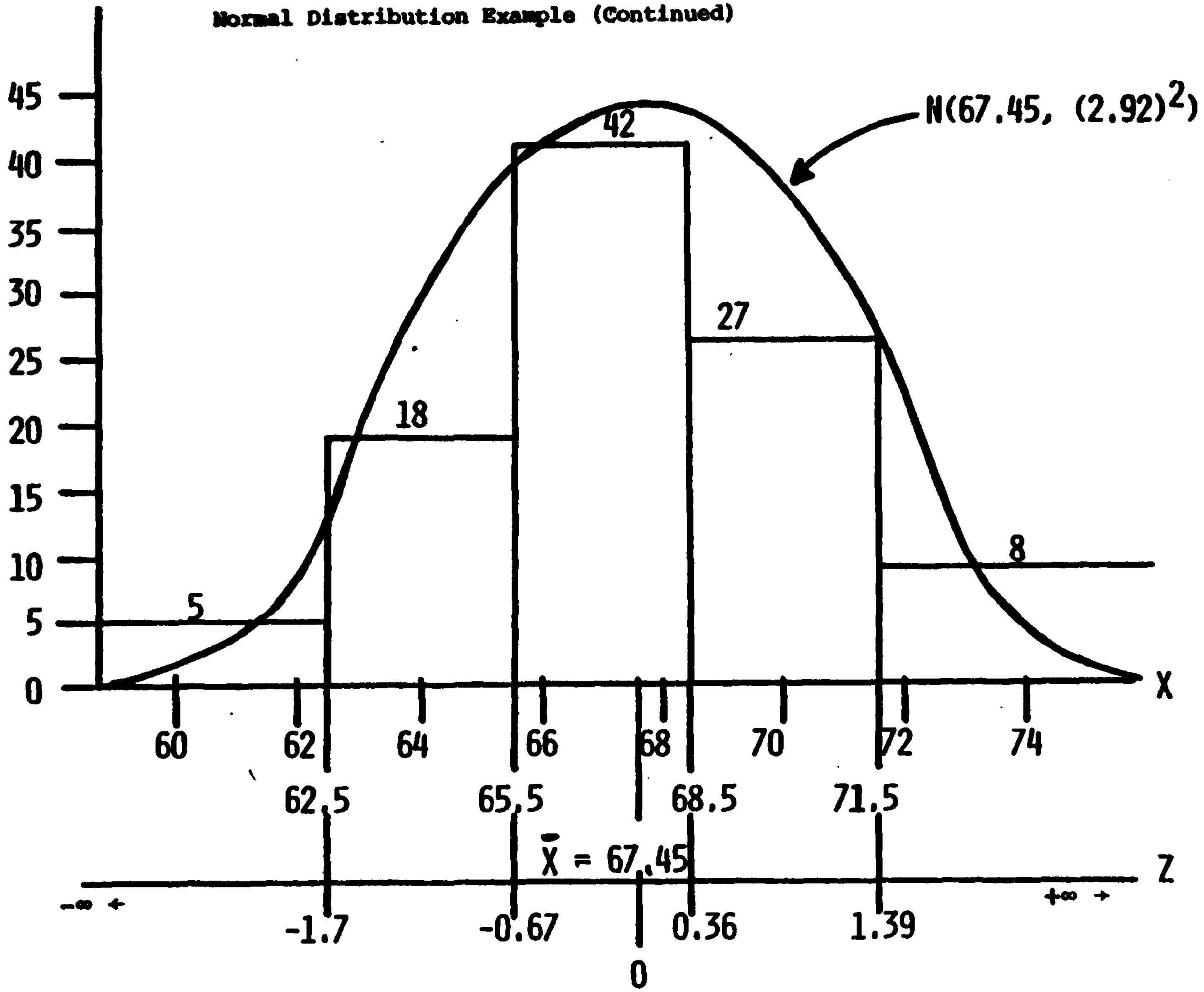
Kolmogorov-Smirnov (K-S)/Lilliefors One-Sample Tests

General description/purpose: The Kolmogorov-Smirnov/Lilliefors one-sample GOF tests may be used to test whether (a) statistically significant difference(s) exist(s) between an empirical distribution and an hypothesized theoretical distribution with a continuous cumulative distribution function, the parameter(s) of which should be pre-specified.

Figure 22-1

X^2 Goodness of fit Test

Normal Distribution Example (Continued)



Assumptions/limitations:

1. The K-S test is designed for the conduct of a goodness of fit test of an empirical or sample distribution against an hypothesized theoretical distribution with a continuous cumulative distribution function (cdf). It may, however, be applied to the question of goodness of fit against an hypothesized theoretical distribution with a discrete cdf. When the cdf is continuous, the level of significance of the test is exactly the pre-specified α-level; when the cdf is discrete, the test is conservative: the true level of significance is at most the pre-specified α-level. (Gibbons: p. 85).

2. The K-S test requires that all parameters of the hypothesized theoretical distribution be specified in the goodness of fit hypothesis test. If any parameters are estimated from the sample data, the exact distribution of the K-S test statistic max $D - D_n$ is no longer known. When K-S $D_{critical}$ values are used, the resulting test is extremely conservative and should be reaccomplished with the Chi-Square test, if at all possible. There are two noteworthy exceptions: when the hypothesized theoretical distribution is normal or exponential, the parameters may be estimated from the sample data, since Lilliefors has tabulated the exact $D_{critical}$ values (obtained by Monte Carlo calculations), Lilliefors' numerical investigations indicate that D_n provides a more powerful test against these distributions than the Chi-Square when his revised $D_{critical}$ values are used. (Gibbons: p. 87).

3. Comparison of K-S and Chi-Square tests:

a. One of the primary advantages of K-S is that the exact distribution of D_n is known and tabulated, whereas the distribution of X_o^2 is only approximately Chi-Square for any finite n. (It is asymptotically X_o^2.)

b. Another advantage of the K-S is that it can be applied for any sample size n, whereas the approximation $X_o^2 \approx X^2$ is only "safe" for large n. Application of the Chi-Square requires satisfaction of the cell size criteria; whenever cell combination is required, the calculated X_o^2 is no longer exact, since it is affected by the scheme of combination. In contradistinction, the K-S treats (ungrouped) individual sample observations separately and, consequently, need not lose information through the combination of cells. The principal advantage of the Chi-Square is that, when the parameters of the hypothesized theoretical distribution are not specified, they can be estimated from the sample data, and a goodness of fit test can be conducted in the usual manner by simply reducing the df of X_o^2.

c. Studies of the power functions of both tests have been reported in the literature in efforts to characterize their relative performances. However, these do not seem to provide a definitive basis for choice, in general.

d. Bottom line: Since the distribution of X_0^2 is known only asymptotically, the Chi-Square test is always approximate; the K-S test is exact against completely specified continuous cdf's and is clearly preferable for small sample sizes. (Gibbons: pp. 86-87.)

Discussion/methodology:

The procedure for application of the Kolmogorov-Smirnov/Lilliefors test involves these steps:

1. Specify the hypothesized cumulative distribution function F(X) including all its parameters, unless a Lilliefors test against a normal or an exponential where the parameter(s) may be estimated from the observations.

2. Arrange the n observations in a sample cumulative distribution (step) function S(X).

3. Pair each interval of S(X) with the comparable interval of F(X).

4. For each step on the cumulative distribution functions, determine the absolute difference(s) between S(X) and F(X).

5. Determine the value of the test statistic, D_n:

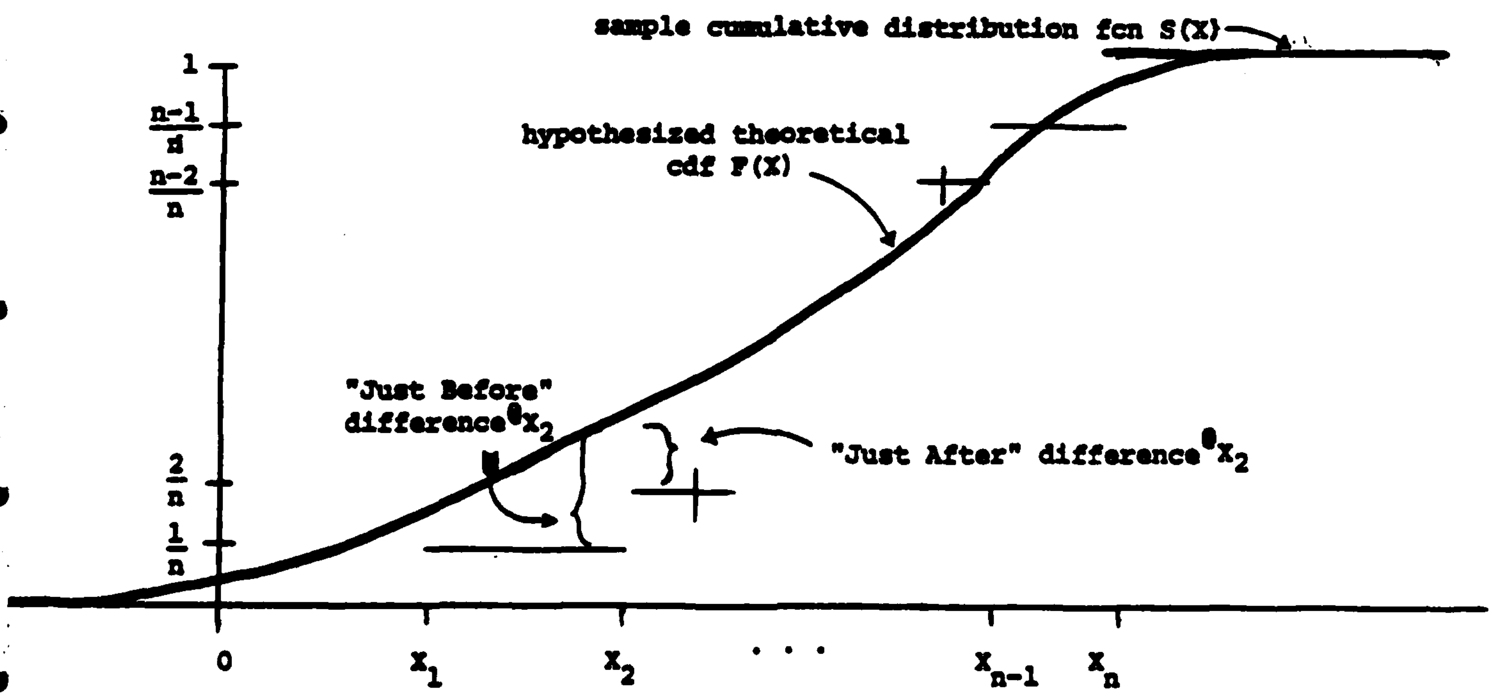
$$D_n = \max D = \max |F(X_i) - S(X_i)| \qquad i = 1, \ldots, n$$

6. Determine from an appropriate table the critical value for the test statistic, $D_{critical}$ or D_c:

$$D_c = D_{\alpha,n}$$

7. Resolve the hypothesis test: if $D_n > D_c$, reject H_o.

Figure 22-2. Construction of K-S/Lilliefors Test Statistics



EXAMPLE 22-14

(Discrete data). A defense contractor contends that the breakaway pressure in lbs. of a certain valve is distributed $N(32, (1.8)^2)$. To test this contention, a random sample of 10 breakaway pressures was obtained as follows: 31.4, 33.5, 31.0, 36-2, 33.3, 33.7, 34.9, 33.4, 37.0. Do these data support the contractor's contention that the breakaway pressure of a certain valve is distributed $N(32, (1.8)^2)$? Test at $\alpha = 0.05$.

x = random variable, breakaway pressure in lbs. of a certain valve.

$H_0 = X \sim N(32, (1.8)^2)$

$H_1 = X \sim N(32, (1.8)^2)$

				Just Before Differences			Just After Differences		
Rank i	Value X_i	Z_i	$F(X_i)$	$\frac{i-1}{n}$	$S(X_i)$	D	$\frac{i}{n}$	$S(X_i)$	D
1	31.0	−.56	.2877	0	0	.2877	1/10	.1	.1877
2	31.4	−.33	.3707	1/10	.1	.2707	2/10	.2	.1707
3	33.3	+.72	.7642	2/10	.2	.5642	3/10	.3	.4642
4	33.4	+.78	.7823	3/10	.3	.4823	4/10	.4	.3823
5	33.5	+.83	.7967	4/10	.4	.3967	5/10	.5	.2967
6	33.7	+.94	.8264	5/10	.5	.3264	6/10	.6	.2264
7	34.4	+1.33	.9082	6/10	.6	.3082	7/10	.7	.2082
8	34.9	+1.6	.9452	7/10	.7	.2452	8/10	.8	.1452
9	36.2	+2.3	.9893	8/10	.8	.1893	9/10	.9	.0893
10	37.0	+2.8	.9974	9/10	.9	.0974	10/10	1.0	.0026

$D_n = D_{10} = \max |F(X_i) - S(X_i)| = 0.5642 \qquad i = 1, \ldots, 10$

$D_c = D_{\alpha,n} = D_{0.05,\,10} = 0.409$ from Table 22-3 (K-S Test Critical Values)

$D_n = 0.5642 > 0.409 = D_c$

Conclusion: Reject H_0. The breakaway pressure of a certain valve is not distributed $N(32, (1.8)^2)$ at the 5% level of significance.

Table 22-3

Critical Values of the One-Sample Kolmogorov-Smirnov Test Statistic: Two-Sided Test

	α=.20	.10	.05	.02	.01
n = 1	.900	.950	.975	.990	.995
2	.684	.776	.842	.900	.929
3	.565	.636	.708	.785	.829
4	.493	.565	.624	.689	.734
5	.447	.509	.563	.627	.669
6	.410	.468	.519	.577	.617
7	.381	.436	.483	.538	.576
8	.358	.410	.454	.507	.542
9	.339	.387	.430	.480	.513
10	.323	.369	.409	.457	.489
11	.308	.352	.391	.437	.468
12	.296	.338	.375	.419	.449
13	.285	.325	.361	.404	.432
14	.275	.314	.349	.390	.418
15	.266	.304	.338	.377	.404
16	.258	.295	.327	.366	.392
17	.250	.286	.318	.355	.381
18	.244	.279	.309	.346	.371
19	.237	.271	.301	.337	.361
20	.232	.265	.294	.329	.352
21	.226	.259	.287	.321	.344
22	.221	.253	.281	.314	.337
23	.216	.247	.275	.307	.330
24	.212	.242	.269	.301	.323
25	.208	.238	.264	.295	.317
26	.204	.233	.259	.290	.311
27	.200	.229	.254	.284	.305
28	.197	.225	.250	.279	.300
29	.193	.221	.246	.275	.295
30	.190	.218	.242	.270	.290
31	.187	.214	.238	.266	.285
32	.184	.211	.234	.262	.281
33	.182	.208	.231	.258	.277
34	.179	.205	.227	.254	.273
35	.177	.202	.224	.251	.269
36	.174	.199	.221	.247	.265
37	.172	.196	.218	.244	.262
38	.170	.194	.215	.241	.258
39	.168	.191	.213	.238	.255
40	.165	.189	.210	.235	.252
	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Table entries are critical values W_α such that Prob $(W \geq W_\alpha) = \alpha$

Source: Adapted from L. H. Miller, "Tables of Percentage Points of Kolmogorov Statistic," Journal of the American Statistical Assoc. 51 (1956), pp. 111-21.

EXAMPLE 22-15

(Blocked data). The normal distribution example first resolved with the Chi square GOF test is now approached as a Lilliefors test application.

X = random variable, the score of an enlistee on a particular aptitude test (integer value)

H_o: $X \sim N(67.45, (2.92)^2)$

H_1: $X \not\sim N(67.45, (2.92)^2)$ $\overline{X} = \hat{\mu}$, $s^2 = \hat{\sigma}^2$, estimated from sample

i	Class Interval	Class Boundray B_i	Z_{B_i}	O_i	$\frac{\Sigma f_i}{n}$	$S(X_i)$	$F(X_i)$	D
		$-\infty$	$-\infty$		0	0	0	0
1	60-62			5				
		62.5	-1.70		$\frac{5}{100}$	0.05	0.0446	0.0054
2	63-65			18				
		65.5	-0.67		$\frac{23}{100}$	0.23	0.2514	0.0214
3	66-68			42				
		68.5	0.36		$\frac{65}{100}$	0.65	0.6406	0.0094
4	69-71			27				
		71.5	1.39		$\frac{92}{100}$	0.92	0.9177	0.0023
5	72-74			8				
		$+\infty$	$+\infty$		$\frac{100}{100}$	1.0	1.0	0
				n = 100				

$$D_n = D_{100} = \max_{\substack{X_i = B_i \\ i = 1, \ldots, 5}} |F(X_i) - S(X_i)| = 0.0214$$

$$D_c = D_{\alpha,n} = D_{0.05,100} = 0.0886$$

from Table 22-4 (Lilliefors normal test critical values)

$$D = 0.0214 < 0.0886 = D_c$$

Conclusion. Fail to reject H_o, etc.

Reference(s)

1. Gibbons, Jean Dickinson. Nonparametric Statistical Inference. (McGraw-Hill: New York, 1971) pp. 73-87.

2. Siegel, pp. 47-52.

Table 22-4

Critical Values of the Lilliefors Test Statistic for an Unspecified Normal Distribution

		α=.20	.15	.10	.05	.01
Sample size n=	4	.300	.319	.352	.381	.417
	5	.285	.299	.315	.337	.405
	6	.265	.277	.294	.319	.364
	7	.247	.258	.276	.300	.348
	8	.233	.244	.261	.285	.331
	9	.223	.233	.249	.271	.311
	10	.215	.224	.239	.258	.294
	11	.206	.217	.230	.249	.284
	12	.199	.212	.223	.242	.275
	13	.190	.202	.214	.234	.268
	14	.183	.194	.207	.227	.261
	15	.177	.187	.201	.220	.257
	16	.173	.182	.195	.213	.250
	17	.169	.177	.189	.206	.245
	18	.166	.173	.184	.200	.239
	19	.163	.169	.179	.195	.235
	20	.160	.166	.174	.190	.231
	25	.149	.153	.165	.180	.203
	30	.131	.136	.144	.161	.187
Over	30	$\frac{.736}{\sqrt{n}}$	$\frac{.768}{\sqrt{n}}$	$\frac{.805}{\sqrt{n}}$	$\frac{.886}{\sqrt{n}}$	$\frac{1.031}{\sqrt{n}}$

Table entries are critical values W_α wuch that Prob $(W \geq W_\alpha) = \alpha$

Source: Adapted from H. W. Lilliefors, "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," Journal of the American Statistical Association, 62 (1967), pp. 399-402.